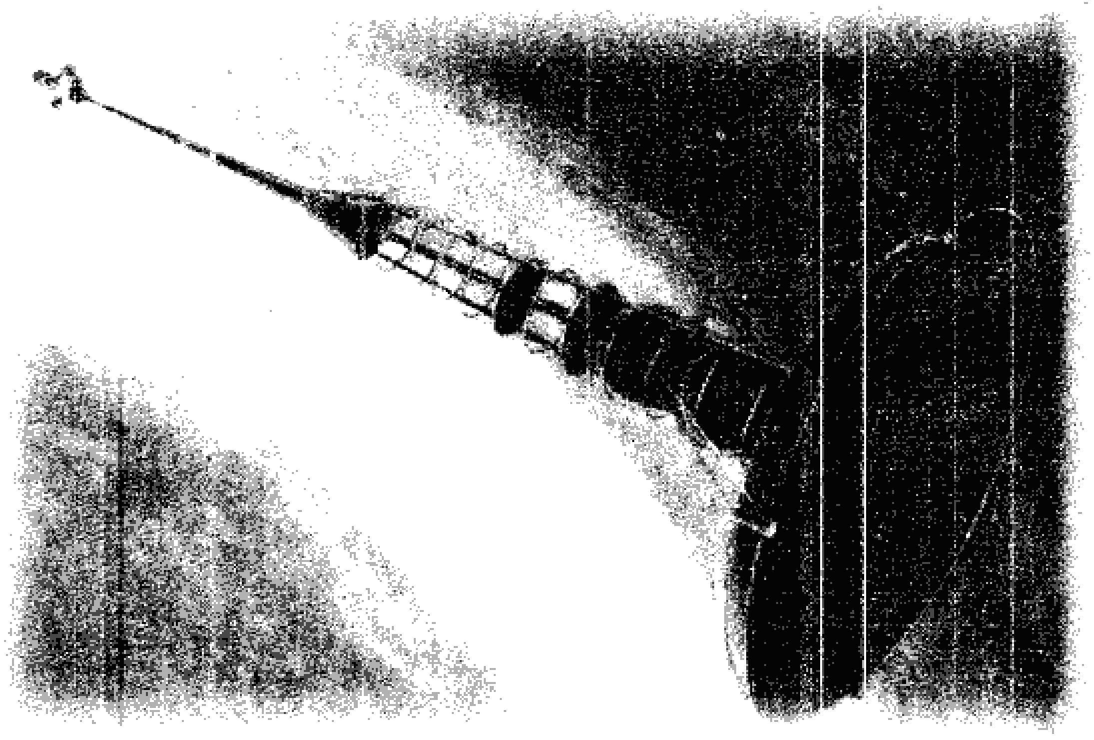
TURING 图灵数学·统计学丛书 31



Design and Analysis of Experiments

实验设计与分析

（第6版）

ouglas C. Montgomery 著

傅珏生 张健 王振羽 解燕 译

人民邮电出版社

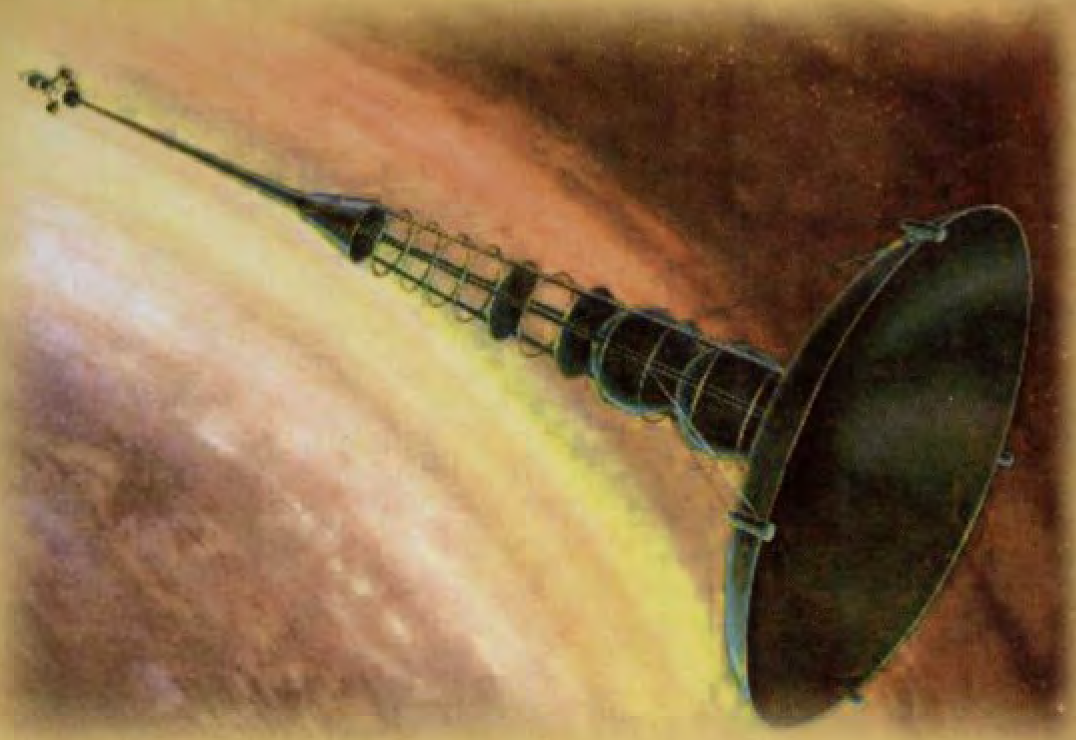
北 京

VISIT…

LANZAROTE Caliente.com

TURING 图灵数学·统计学丛书 31

WILEY



Design and Analysis of Experiments

实验设计与分析

（第 6 版）

[美] Douglas C. Montgomery 著

傅珏生 张健 王振羽 解燕 译

人民邮电出版社
POSTS & TELECOM PRESS

实验设计与分析（第6版）
Design and Analysis of Experiments

“实验设计的经典教材和实用参考书，学生们会发现，在上完课进入到实际工作中以后，你还需要经常参考书中的相关理论、方法和技巧。”

——*Quality & Reliability Engineering International*杂志

“这是实验设计领域最权威的一本书，全面、透彻而且非常实用。”

——Michael R. Chernick，著名统计学家，*Bootstrap Methods*一书作者

本书是实验设计与分析课程的经典教材，凝聚了作者在美国多所名校近40年的教学经验，同时充分展现了作者将统计实验方法应用于各行各业实际项目的丰富工程实践经验。已被美国麻省理工学院、普度大学、华盛顿大学、英国曼彻斯特大学和中国台湾大学等世界众多高校广泛采用，并深受广大工程科技人员的欢迎。原版销量累计已经超过10万册。

本书讲述了设计、实施和分析实验以改善产品与过程性能的高效方法，说明了如何使用统计实验进行产品的设计与开发、改进生产过程、获取系统特征和优化的信息。

本书特色

- 注重与信息技术相结合，给出了Design-Expert和Minitab两种软件的输出结果。
- 内容全面，实例丰富，选材新颖。
- 本书网站提供了大量的支持性资源和补充材料，有利于读者深入学习。

Douglas C. Montgomery 著名统计学家，亚利桑那州立大学工业与管理系统工程教授，美国统计学会、工业工程学会、质量控制学会会士。他出版了多部影响深远的统计学著作，发表了大量广为引用的论文。他还应IBM、可口可乐、波音、摩托罗拉等著名公司邀请开展合作项目，在半导体、医疗设备、生物技术等领域深入地进行统计方法的实践研究和理论探讨。

WILEY
www.wiley.com
本书相关信息请访问：图灵网站 http://www.turingbook.com
读者热线：(010)88593802
反馈/投稿/推荐信箱：contact@turingbook.com
分类建议 统计学/实验设计
人民邮电出版社网址 www.ptpress.com.cn
ISBN 978-7-115-19234-9
9 787115 192349 >
ISBN 978-7-115-19234-9/O1
定价：89.00 元

版 权 声 明

Design and Analysis of Experiments, 6th Edition by Douglas C. Montgomery, ISBN 0-471-48735-X.

Copyright © 2005 John Wiley & Sons, Inc. All rights reserved.

Authorized translation of the edition published by John Wiley & Sons, New York, Chichester, Brisbane, Singapore and Toronto. No part of this book may be reproduced in any form without the written permission of John Wiley & Sons, Inc.

This edition published by POSTS & TELECOM PRESS Copyright © 2009.

本书简体中文版由 John Wiley & Sons, Inc.授权人民邮电出版社在全球出版发行. 未经出版者书面许可, 不得以任何方式复制本书内容.

本书封底贴有 John Wiley & Sons, Inc.激光防伪标签, 无标签者不得销售.

版权所有, 侵权必究.

译者简介

傅珏生 男, 江苏苏州人. 1962 年 9 月生, 2003 年毕业于香港中文大学并获博士学位. 苏州大学数学科学学院统计系副教授, 主要研究方向是实验设计、贝叶斯统计. 现担任中国现场统计研究会理事、苏州市现场统计研究会理事长、《数理统计与管理》编委. 作为主要参加者参加了国家自然基金项目等科研项目.

张　健 男, 江苏昆山人. 1972 年出生, 2004 年毕业于苏州大学获理学博士学位. 苏州大学数学科学学院讲师, 主要研究方向是实验设计. 作为主要参加者参加了国家自然基金项目等科研项目.

王振羽 女, 江苏苏州人. 1957 年 6 月生, 1986 年华东师范大学数理统计专业研究生毕业. 苏州大学数学科学学院统计系副教授, 主要研究方向是实验设计、质量管理中的统计方法等. 现担任江苏省现场统计研究会理事、苏州市现场统计研究会副理事长. 作为主要参加者参加了国家自然基金项目等科研项目, 在多家中外企业协作开展科研协作工作等.

解　燕 女, 江苏海安人. 1968 年 2 月出生, 2003 年毕业于苏州大学并获硕士学位. 苏州大学医学部从事学生管理工作, 主要研究方向是抽样调查、实验设计. 现担任江苏省现场统计研究会理事、苏州市现场统计研究会秘书长.

译 者 序

本书是关于实验设计与分析的名著, 已经是第 6 版了. 作者是亚利桑那州立大学工业与管理系统工程方面的教授, 他是一位杰出的应用统计学者, 多次获得国际性学术奖励.

本书强调了工程师和科学工作者在产品设计和开发、工序开发及改进等方面用实验设计作为工具的重要性, 案例丰富, 所有重要的概念、方法都通过大量实例予以说明.

与前 5 版相比, 作者进一步强调了实验设计的应用背景, 给出了实际应用方面的一些新的进展与实际案例. 本书还强调统计软件 Minitab 和 Design-Expert 的使用.

本书的翻译工作由傅珏生主持实施. 解燕、傅珏生、王振羽和张健分别完成了原著第 1 章至第 4 章、第 5 章至第 8 章、第 9 章至第 12 章和第 13 章至第 15 章的翻译初稿, 并对所有案例重新进行计算, 纠正了原著中的一些错误. 4 位译者共同校对与修改, 最后由傅珏生协调定稿.

译者在翻译过程中始终得到了第 3 版译者汪仁官、陈荣昭两位先生的帮助, 在此表示感谢. 同时感谢人民邮电出版社图灵公司的领导和编辑在本书的翻译和出版过程中给予的帮助和鼓励. 感谢苏州大学数学科学学院的大力支持.

由于我们水平有限, 译文中难免有不当或错讹之处, 敬请读者批评指正.

译 者

2008 年 8 月

前 言

读者对象

这是一本论述实验设计与分析的入门教科书. 它是基于我 40 年来在亚利桑那州立大学、华盛顿大学和佐治亚理工学院讲授实验设计方面的大学课程而写成的. 它还包括了我在专业实践中认为有用的实验设计方法, 多年来我在自然科学和工程等诸多领域内担任统计顾问, 并涉足产品实现过程.

本书供学完统计方法基础课程的读者使用. 读者至少应掌握描述性统计技术、正态分布、置信区间以及关于均值和方差的假设检验的基本知识和有关概念. 第 10 章至第 12 章的部分内容要求读者熟知矩阵代数.

因为所需的预备知识相对适中, 所以本书可以作为工程学、物理学、数学、化学以及其他学科本科生的侧重于实验统计设计的后继统计学教程. 多年来我将本书作为工科研究生一年级的教材, 选修这门课的学生的专业背景有工程学、材料学、物理学、化学、数学、运筹学和统计学等. 我还将它作为各种背景的技术员的短期技术培训教材. 书中的大量例子说明了所有的设计和分析方法, 这些例子都体现了实验设计在现实世界中的应用, 并选自不同的工程学及科学领域. 这为培养工程师和科学家的理论课程平添了浓郁的应用气息, 也使本书可用作各种学科的实验者的参考书.

关于本书

第 6 版是本书的重要修订版. 在努力保持以前版本的设计和分析这两方面内容平衡的同时, 我添加并组织了很多新的内容和例子. 本版更加强调计算机的应用.

1. Minitab 软件和 Design-Expert 软件

近几年来, 在实验设计和分析领域出现了一批优秀的软件产品. 本书多处使用了其中两种软件 (Minitab 和 Design-Expert) 的输出结果. Minitab 是一个应用广泛的统计软件包, 它有良好的数据分析能力和相当好的处理固定因子及随机因子 (包括混合模型) 的实验分析能力. Design-Expert 是一个侧重于实验设计的软件包, 它具有构建和评估设计的能力并可以进行深入的分析. 本书的例子都可以用 Design-Expert 和 Minitab 的学生版完成, 并强烈推荐使用. 我极力主张使用本书的教师在课程教学中加入计算机软件. (在我的课程中, 每一讲都用便携式计算机和投影仪, 在课堂中讨论的每一个设计或分析的主题我都用计算机演示过.)

2. 经验模型

我一如既往地注重实验本身和实验者通过该实验的结果建立的模型之间的联系. 工程师 (乃至广义上的物理学家、化学家) 在他们早期的科学训练中学习了物理知识及机理模型, 并在他们的大部分职业生涯中都自觉不自觉地运用了这些模型. 统计实验设计为工程师开发一个系统的经验模型提供了理据方面的基础. 可以像使用其他的工程模型一样来使用这个经验模型 (也许借助于响应曲面或等高线, 也许借助于数学方法). 通过多年的教学, 我得出这样一个结论: 运用统计实验设计可以非常有效地调动工程师的创造热情. 因此, 实验的基础经验模型和响应曲面的概念出现在早期版本中并得到了进一步的重视.

3. 析因设计

我也作了一些努力, 尽快将读者带入包括析因设计在内的许多重要论题. 为此, 我将完全随机化单因子实验设计的介绍性材料和方差分析的内容压缩到一章里 (第 3 章). 之所以扩充析因设计和分式析因设计 (第 5 章至第 9 章) 的内容, 目的是为了兼顾一般读者和教师两种读者, 同时将更多的重点放在经验模型上. 许多重要的论题包含了一些新的内容, 包括分式析因后的追加试验, 还有小规模、有效分辨度为Ⅳ和Ⅴ的设计.

4. 增加的重要论题

响应曲面这一章 (第 11 章) 紧接在析因设计、分式析因设计和回归模型之后. 另增加了新的一章 (第 12 章), 其内容为稳健参数设计和工序稳健设计. 第 13 章和第 14 章主要讨论了随机效应实验及其概念在嵌套设计和裂区设计的实际应用. 因为工业界对这些设计越来越感兴趣, 第 13 章和第 14 章还介绍了一些新的论题. 第 15 章概述了一些重要的设计和分析论题: 非正态响应, 选择变换形式的 Box-Cox 方法以及其他的替代方法; 不平衡析因实验; 协方差分析, 包括协变量的析因设计和重复测量.

5. 实验设计

通贯本书, 我始终强调工程师和科学工作者在产品设计和开发、工序开发和改进等方面用实验设计作为工具的重要性, 并说明了利用实验设计有助于工程师开发出不受环境因子和其他变异来源影响的所谓稳健产品. 我相信, 在产品工序开发的早期阶段, 成功地应用实验设计会大大缩短开发时间和降低成本, 而且比起用其他方法来, 所开发的工序和产品会有更好的性能, 并有更高的可靠性.

本书内容很难在通常的一门课程中完成, 我希望教师能针对不同的课程灵活地选择内容, 或者根据师生的兴趣更为深入地讨论某些论题. 每章之末都有一些思考题 (第 1 章除外), 这些思考题的计算量有大有小, 它们有助于加强基础理论的学习, 以及基本原理的推广或完善.

课程建议

我自己的课程着重于析因设计和分式析因设计, 因此, 我通常选择第 1 章、第 2 章 (非常快)、第 3 章的大部分, 以及第 4 章 (不完全区组的材料除外, 并仅简单介绍拉丁方). 我会详细讨论第 5 章至第 8 章的关于析因设计和二水平的分式析因设计. 课程最后引进响应曲面方法 (第 11 章), 并简要介绍随机效应模型 (第 13 章) 以及嵌套设计和裂区设计 (第 14 章). 我通常要求学生完成学期作业, 包括设计、运行并给出统计实验设计的结果. 我要求学生以小组的形式完成学期作业, 因为大部分工业实验都是这样进行的. 他们必须给出作业的结果, 包括口头形式和书面形式.

补充材料

第 6 版为每一章都准备了一些补充材料. 本书无法详尽讨论的论题, 以及一些没有直接放在本书中的、但它们的引入对于一些学生和专业人员来说是非常有用的主题都放在补充材料里. 部分材料超出了本书的数学水平. 拥有大批听众的教师和一些更高级的设计课程可能会从一些补充材料的论题中受益. 登录网页 http://tinyurl.com/Montgomery Experiment 可以获得这些

材料的电子版.

网页

教师和学生可以访问网页http://tinyurl.com/Montgomery Experiment, 获得支持性材料, 为了更有效的使用本书, 我们将这个网页作为交流创新思想和建议信息的平台. 访问该网页, 可以得到前面所描述的补充材料, 以及各种例子、家庭作业以及亚利桑那州立大学部分学生该课程的优秀学期项目的电子版本.

1. 学生社区

课本网页的学生部分包括以下内容:

(1) 前面提及的补充材料;

(2) 本书各种例题和家庭作业问题 (电子形式);

(3) 学生专题范本.

2. 教师社区

课本网站的教师部分包括以下内容:

(1) 本书习题解答;

(2) 前面提及的补充材料;

(3) PowerPoint 讲稿幻灯片;

(4) 本书各种例题和家庭作业问题 (电子形式);

(5) 学生专题范本.

部分内容仅提供给使用本书作为教材的教师, 有密码保护. 可致信 contact@turingbook.com 申请密码.

*3. 学生解题手册

学生解题手册的目的是帮助学生更深程度地理解如何运用本书提及到的概念. 通过详细地讲解如何解答精心选择的习题, 可以培养实际应用的敏锐眼光.

所提供的解答针对的是本人所选择的问题. 偶尔给出了一些 "延伸练习" 并向学生提供特定数据集的详尽解答. 收录在学生解题手册中的题目, 在本书思考题题号的左上角以 "*" 标示.

该手册是一本非常好的学习参考资料, 大多数读者都会觉得非常有用. 可以和本书一起定购, 也可以单独购买. 联系当地的 Wiley 代表从书店或者登录 Wiley 网站购买.

致谢

感谢用过本书前 5 版的教师、学生和同行, 他们对本书的再版提出了富有建设性的建议. Raymond H. Myers 博士、G. Geoffrey Vining 博士、Dennis Lin 博士、John Ramberg 博士、Joseph Pignatiello 博士、Lloyd S. Nelson 博士、Andre Khuri 博士、Peter Nelson 博士、John A. Cornell 博士、George C. Runger 博士、Bert Keats 博士、Dwayne Rollier 博士、Norma Hubele 博士、Murat Kulahci 博士、Cynthia Lowry 博士、Russell G. Heikes 博士、Harrison M. Wadsworth 博士、William W. Hines 博士、Arvind Shah 博士、Jane Ammons 博士、Diane Schaub 博士、Mark Anderson 先生、Pat Whitcomb 先生、Pat Spagon

博士、William DuMouche 博士的建议尤其宝贵. 我的系主任 Gary Hogg 博士为我提供了激发智慧的工作环境.

感谢和我一起工作过的专业实践人员的帮助. 这里不能把每一位都提到了, 其中最主要的有: 敏迅 (Mindspeed) 公司的 Dan McCarville 博士, 乔治集团 (George Group) 的 Lisa Custer 博士, 英特尔公司的 Richard Post 博士, 波音公司的 Tom Bingham 先生、Dick Vaughn 先生、Julian Anderson 博士、Richard Alkire 先生以及 Chase Neilson 先生, 美铝 (Alcoa) 公司的 Mike Goza 先生、Don Walton 先生、Karen Madison 女士、Jeff Stevens 先生以及 Bob Kohm 先生, IBM 公司的 Jay Gardiner 博士、John Butora 先生、Dana Lesher 先生、Lolly Marwah 先生以及 Leon Mason 先生, Somatech 公司的 Paul Tobias 博士, 可口可乐公司的 Elizabeth A. Peck 女士, Signetics 公司的 Sadri Khalessi 博士和 Franz Wagner 先生, Monsanto 化学制品公司的 Robert V. Baxley 先生, 精密零件铸造 (Precision Castparts) 公司的 Harry Peterson-Nedry 先生和 Russell Boyles 博士, 联合信号航空航天 (Allied-Signal Aerospace) 公司的 Bill New 先生和 Randy Schmid 先生, John Fluke 制造公司的小 John M. Fluke 先生, 佐治亚太平洋 (Georgia-Pacific) 公司的 Larry Newton 先生和 Kip Howlett 先生, BBN 软件产品公司的 Ernesto Ramos 博士.

我也非常感谢 E. S. Pearson 教授和 *Biometrika* 杂志的理事们, 以及 John Wiley & Sons 公司、Prentice-Hall 公司、美国统计学会、数理统计研究所, 还有允许我们复制版权资料的 *Biometrics* 杂志的编者. Lisa Custer 博士为教师解题手册做出了极好的准备工作, Cheryl Jennings 博士和 Sarah Streett 博士提供了迅捷高效的校对援助. 还要感谢美国海军研究局、国家科学基金会 (National Science Foundation, NSF)、NSF/Industry/UCRC(基金会/企业/高校合作研究中心) 在亚利桑那州立大学的质量和可靠性工程研究中心的各会员公司和 IBM 公司, 感谢它们对我在工程统计学和实验设计方面的研究工作给予的支持.

Douglas C. Montgomery

亚利桑那州坦佩市

图书在版编目(CIP)数据

实验设计与分析：第6版 /(美)蒙哥马利(Montgomery, D. C.)著；傅珏生等译. —北京：人民邮电出版社，2009. 1
(图灵数学·统计学丛书)
书名原文：Design and Analysis of Experiments, 6th Edition
ISBN 978-7-115-19234-9
Ⅰ.实… Ⅱ.①蒙… ②傅… Ⅲ.试验设计(数学)－高等学校－教材 Ⅳ.O212.6
中国版本图书馆 CIP 数据核字(2008)第181473号

内 容 提 要

本书作为实验设计与分析领域的名著，是作者在亚利桑那州立大学、华盛顿大学和佐治亚理工学院三所大学近40年实验设计教学经验的基础上编写的。全书内容广泛，实例丰富，包括简单比较试验、析因设计、分式析因设计、拟合回归模型、响应曲面方法和设计、稳健参数设计和过程稳健性研究、含随机因子的实验、嵌套设计和裂区设计等。

本书可作为自然科学研究人员、工程技术人员、管理人员进行科学实验设计与分析的参考书，也可作为农林类、医学类、生物类、统计类的教师和高年级本科生和研究生的教学参考用书。

图灵数学·统计学丛书

实验设计与分析（第6版）

◆ 著　[美] Douglas C. Montgomery
译　傅珏生　张　健　王振羽　解　燕
责任编辑　明永玲
执行编辑　张绍发

◆ 人民邮电出版社出版发行　北京市崇文区夕照寺街14号
邮编　100061　电子函件　315@ptpress.com.cn
网址：http://www.ptpress.com.cn
北京铭成印刷有限公司印刷

◆ 开本：700×1000　1/16
印张：35.25
字数：882千字　2009年1月第1版
印数：1–3 000 册　2009年1月北京第1次印刷

著作权合同登记号　图字：01-2006-5787号

ISBN 978-7-115-19234-9/O1

定价：89.00元

读者服务热线：(010)88593802　印装质量热线：(010) 67129223
反盗版热线：(010)67171154

目　录

第 1 章　引言 ······ 1
1.1 实验策略 ······ 1
1.2 实验设计的一些典型应用 ······ 6
1.3 基本原理 ······ 8
1.4 设计实验指南 ······ 10
1.5 统计设计简史 ······ 14
1.6 小结: 在实验中使用统计方法 ······ 16
1.7 思考题 ······ 16
第 2 章　简单比较实验 ······ 17
2.1 引言 ······ 17
2.2 基本统计概念 ······ 18
2.3 抽样与抽样分布 ······ 21
2.4 关于均值差的推断, 随机化设计 ······ 26
2.4.1 假设检验 ······ 26
2.4.2 样本量的选取 ······ 32
2.4.3 置信区间 ······ 34
2.4.4 $\sigma_1^2 \neq \sigma_2^2$ 的情形 ······ 36
2.4.5 σ_1^2 与 σ_2^2 为已知的情形 ······ 36
2.4.6 均值与已知值的比较 ······ 37
2.4.7 小结 ······ 37
2.5 关于均值差的推断, 配对比较设计 ······ 38
2.5.1 配对比较问题 ······ 38
2.5.2 配对比较设计的优点 ······ 40
2.6 正态分布方差的推断 ······ 42
2.7 思考题 ······ 43
第 3 章　单因子实验: 方差分析 ······ 48
3.1 一个例子 ······ 48
3.2 方差分析 ······ 50
3.3 固定效应模型的分析 ······ 52
3.3.1 总平方和的分解 ······ 53
3.3.2 统计分析 ······ 55
3.3.3 模型参数的估计 ······ 59
3.3.4 不平衡数据 ······ 60
3.4 模型适合性检验 ······ 60
3.4.1 正态性假设 ······ 61
3.4.2 依时间序列的残差图 ······ 62
3.4.3 残差与拟合值的关系图 ······ 63
3.4.4 残差与其他变量的关系图 ······ 67
3.5 结果的实际解释 ······ 68
3.5.1 回归模型 ······ 68
3.5.2 处理均值的比较 ······ 70
3.5.3 均值的图解比较法 ······ 70
3.5.4 对照法 ······ 71
3.5.5 正交对照法 ······ 73
3.5.6 用来比较全部对照的 Scheffé方法 ······ 74
3.5.7 处理均值的配对比较法 ······ 76
3.5.8 将各个处理均值与一个控制进行比较 ······ 78
3.6 计算机输出示例 ······ 79
3.7 确定样本量 ······ 82
3.7.1 抽检特性曲线 ······ 82
3.7.2 规定标准差的增量 ······ 84
3.7.3 置信区间的估计方法 ······ 85
3.8 寻找分散效应 ······ 85
3.9 方差分析的回归处理法 ······ 87
3.9.1 模型参数的最小二乘估计 ······ 87
3.9.2 一般回归显著性检验 ······ 88
3.10 方差分析中的非参数方法 ······ 90
3.10.1 Kruskal-Wallis 检验法 ······ 90
3.10.2 关于秩变换的一般评论 ······ 91
3.11 思考题 ······ 91
第 4 章　随机化区组, 拉丁方, 以及有关的设计 ······ 98
4.1 随机化完全区组设计 ······ 98

4.1.1 RCBD 的统计分析 ········ 99
4.1.2 模型适合性检验 ·········· 105
4.1.3 随机化完全区组设计的一些其他方面 ············ 107
4.1.4 估计模型参数和一般回归显著性检验 ·············· 109
4.2 拉丁方设计 ······················ 112
4.3 正交拉丁方设计 ·················· 117
4.4 平衡不完全区组设计 ············· 119
4.4.1 BIBD 的统计分析 ········ 120
4.4.2 参数的最小二乘估计 ······ 125
4.4.3 BIBD 中内部信息的恢复 ···················· 125
4.5 思考题 ························ 128
第 5 章 析因设计导引 ················ 134
5.1 基本定义与原理 ················ 134
5.2 析因设计的优点 ················ 137
5.3 二因子析因设计 ················ 138
5.3.1 一个例子 ················ 138
5.3.2 固定效应模型的统计分析 ···················· 140
5.3.3 模型适合性检验 ·········· 146
5.3.4 估计模型参数 ············ 147
5.3.5 样本量的选择 ············ 148
5.3.6 假定在二因子模型中没有交互作用 ················ 149
5.3.7 每单元一个观测值 ········ 150
5.4 一般的析因设计 ················ 152
5.5 拟合响应曲线与曲面 ············ 157
5.6 析因设计中的区组化 ············ 161
5.7 思考题 ························ 165
第 6 章 2^k 析因设计 ················ 171
6.1 引言 ························ 171
6.2 2^2 设计 ······················ 171
6.3 2^3 设计 ······················ 178
6.4 一般的 2^k 设计 ················ 189
6.5 2^k 设计的单次重复 ············ 191
6.6 附加中心点的 2^k 设计 ·········· 208
6.7 使用规范化设计变量的理由 ········ 212
6.8 思考题 ························ 214
第 7 章 2^k 析因实验的区组设计和混区设计 ···················· 225
7.1 引言 ························ 225
7.2 重复的 2^k 析因实验的区组设计 ···· 225
7.3 2^k 析因实验的混区设计 ·········· 226
7.4 二区组的 2^k 析因实验的混区设计 ························ 226
7.5 区组化重要性的另一个例证 ········ 231
7.6 四区组的 2^k 析因实验的混区设计 ························ 233
7.7 2^p 个区组的 2^k 析因实验的混区设计 ························ 234
7.8 部分混区设计 ·················· 236
7.9 思考题 ························ 238
第 8 章 二水平分式析因设计 ·········· 239
8.1 引言 ························ 239
8.2 2^k 析因设计的 $\frac{1}{2}$ 分式设计 ········ 240
8.2.1 定义与基本原理 ·········· 240
8.2.2 设计分辨度 ·············· 242
8.2.3 $\frac{1}{2}$ 分式设计的构造与分析 ·················· 242
8.3 2^k 析因设计的 $\frac{1}{4}$ 分式设计 ········ 251
8.4 一般的 2^{k-p} 分式析因设计 ········ 257
8.4.1 设计的选择 ·············· 257
8.4.2 2^{k-p} 分式析因设计的分析 ···················· 259
8.4.3 分式析因设计的区组化 ····· 260
8.5 分辨度为III的设计 ·············· 264
8.5.1 分辨度为III的设计的构造 ···················· 264
8.5.2 折叠分辨度为III的分式设计以分离别名效应 ······· 266
8.5.3 Plackett–Burman 设计 ···· 269
8.6 分辨度为IV和V的设计 ··········· 272
8.6.1 分辨度为IV的设计 ········· 272
8.6.2 分辨度为IV的设计的序贯实验 ···················· 275

8.6.3 分辨度为V的设计 ········· 281
8.7 超饱和设计 ········· 283
8.8 小结 ········· 284
8.9 思考题 ········· 284
第 9 章 三水平和混合水平析因设计与分式析因设计 ········· 296
9.1 3^k 析因设计 ········· 296
9.1.1 3^k 设计的记号和引进动机 ········· 296
9.1.2 3^2 设计 ········· 297
9.1.3 3^3 设计 ········· 299
9.1.4 一般的 3^k 设计 ········· 302
9.2 3^k 析因设计的混区设计 ········· 303
9.2.1 三区组的 3^k 析因设计 ········· 303
9.2.2 九区组的 3^k 析因设计 ········· 306
9.2.3 3^p 个区组的 3^k 析因设计 ········· 307
9.3 3^k 析因设计的分式重复 ········· 308
9.3.1 3^k 析因设计的 $\frac{1}{3}$ 分式设计 ········· 308
9.3.2 其他的 3^{k-p} 分式析因设计 ········· 310
9.4 混合水平的析因设计 ········· 312
9.4.1 二水平和三水平的因子 ········· 312
9.4.2 二水平和四水平的因子 ········· 313
9.5 思考题 ········· 314
第 10 章 拟合回归模型 ········· 319
10.1 引言 ········· 319
10.2 线性回归模型 ········· 319
10.3 线性回归模型的参数估计 ········· 320
10.4 多元回归的假设检验 ········· 332
10.4.1 回归的显著性检验 ········· 332
10.4.2 回归系数的个别检验和分组检验 ········· 334
10.5 多元回归的置信区间 ········· 337
10.5.1 单个回归系数的置信区间 ········· 337
10.5.2 平均响应的置信区间 ········· 337
10.6 新响应观测的预测 ········· 338
10.7 回归模型的诊断 ········· 339
10.7.1 尺度残差和 PRESS ········· 339
10.7.2 影响诊断 ········· 341
10.8 拟合不足检验 ········· 342
10.9 思考题 ········· 344
第 11 章 响应曲面法与设计 ········· 347
11.1 响应曲面法引言 ········· 347
11.2 最速上升法 ········· 349
11.3 二阶响应曲面的分析 ········· 354
11.3.1 稳定点的位置 ········· 354
11.3.2 响应曲面的刻画 ········· 356
11.3.3 岭系统 ········· 362
11.3.4 多重响应 ········· 362
11.4 拟合响应曲面的实验设计 ········· 366
11.4.1 拟合一阶模型的设计 ········· 367
11.4.2 拟合二阶模型的设计 ········· 367
11.4.3 响应曲面设计的区组化 ········· 373
11.4.4 计算机生成 (最优) 设计 ········· 376
11.5 混料实验 ········· 380
11.6 调优运算 ········· 388
11.7 思考题 ········· 392
第 12 章 稳健参数设计与过程稳健性研究 ········· 398
12.1 引言 ········· 398
12.2 直积表设计 ········· 399
12.3 直积表设计的分析 ········· 401
12.4 组合表设计及响应模型法 ········· 403
12.5 设计的选择 ········· 408
12.6 思考题 ········· 411
第 13 章 含随机因子的实验 ········· 415
13.1 随机效应模型 ········· 415
13.2 含随机因子的二因子析因设计 ········· 420
13.3 二因子混合模型 ········· 423
13.4 含随机效应的样本量的确定 ········· 428
13.5 期望均方的计算法则 ········· 429
13.6 近似 F 检验 ········· 432

13.7 关于方差分量估计的一些其他论题 ··· 438
13.7.1 方差分量的近似置信区间 ··· 438
13.7.2 修正的大样本方法 ··· 440
13.7.3 方差分量的最大似然估计 ··· 441
13.8 思考题 ··· 445
第 14 章 嵌套设计和裂区设计 ··· 450
14.1 二级嵌套设计 ··· 450
14.1.1 统计分析 ··· 451
14.1.2 诊断检测 ··· 455
14.1.3 方差分量 ··· 456
14.1.4 交错嵌套设计 ··· 457
14.2 一般的 m 级嵌套设计 ··· 457
14.3 含被套因子和交叉因子的设计 ··· 459
14.4 裂区设计 ··· 463
14.5 裂区设计的其他变形 ··· 468
14.5.1 多于两个因子的裂区设计 ··· 468
14.5.2 裂裂区设计 ··· 471
14.5.3 带裂区设计 ··· 473
14.6 思考题 ··· 474
第 15 章 其他设计与分析论题 ··· 480
15.1 非正态响应与变换 ··· 480
15.1.1 选择一个变换：Box-Cox 方法 ··· 480
15.1.2 广义线性模型 ··· 482
15.2 析因设计中的不平衡数据 ··· 489
15.2.1 成比例的数据：简单的情况 ··· 490
15.2.2 近似方法 ··· 491
15.2.3 精确方法 ··· 492
15.3 协方差分析 ··· 493
15.3.1 方法说明 ··· 494
15.3.2 计算机输出 ··· 500
15.3.3 用一般线性回归显著性检验进行推导 ··· 501
15.3.4 含协变量的析因实验 ··· 503
15.4 重复测量 ··· 507
15.5 思考题 ··· 509
参考文献(图灵网站下载)
附录 ··· 512
索引 ··· 545

第 1 章　引　　言

本章纲要

1.1　实验策略
1.2　实验设计的一些典型应用
1.3　基本原理
1.4　设计实验指南
1.5　统计设计简史
1.6　小结：在实验中使用统计方法
第 1 章补充材料
S1.1　设计实验的更多内容
S1.2　实验设计前的空白指导图
S1.3　实验设计的 Montgomery 理论

1.1　实验策略

研究者几乎在所有的研究领域中都会进行实验，通常是为了探究一个特定过程或系统. 从字面意义上说，一次实验就是一次**试验**(test). 更正式的说法是，我们可以定义一次**实验**(experiment)是一次试验或一系列试验，在实验中通过对某一过程或系统的输入变量作一些有目的的改变，以便能够观测到和识别出可在输出响应中观测到的变化的缘由.

本书讲解实验的设计、实施以及结果数据分析，以便能够得到有效且客观的结论. 我们重点关注工程学和科学方面的实验. 实验在包括新产品设计和配方、制造过程 (process) 开发以及过程改进在内的**产品实现**(product realization) 中起着重要的作用. 多数情况下是为了开发一种**稳健的**(robust) 过程，即受外部变异性来源影响最小的过程.

举一个实验的例子，假设一位冶金工程师想要研究两种不同的淬火工艺 (油淬和盐水淬) 对一种铝合金的效应. 在这里，实验者的目的是确定哪种淬火溶液能使得这种特殊合金的硬度最大. 工程师决定对每种淬火介质都提供若干合金试件并在淬火后测量试件的硬度. 以试件在每种淬火溶液中处理之后的平均硬度来确定哪一种溶液是最好的.

考虑这样一项简单的实验时，有很多重要的问题需要思考：

(1) 要研究的淬火介质只有这两种溶液吗？

(2) 在这项实验中，有没有其他可能影响硬度的因子应该被研究或被控制呢 (比如，淬火介质的温度)？

(3) 每种淬火溶液中应该检测多少块合金试件呢？

(4) 怎样把试件分配给淬火溶液？又应该按什么次序来收集数据？

(5) 应该用什么样的数据分析方法？

(6) 两种淬火介质间的平均观测硬度之差是多大时将被认为是重要的？

所有这些问题，也许还有很多其他问题，都必须在进行这项实验之前得到满意的回答.

实验是**科学**(或**工程**)**方法**的一项重要组成部分. 有时，由于对科学现象理解得比较透彻，直接应用这些简单的原理就能得到包括数学模型的有用结果. 直接由物理原理得到的模型叫**机理模型**(mechanistic model). 一个简单的例子就是电路中关于电流的著名公式，即欧姆定律：$E = IR$.

然而, 科学和工程的大多数问题要求对系统的运行进行**观测**(observation) 和**实验**, 从而阐明系统运行的原因和方式. 设计良好的实验常常可以导出系统运行的模型, 这种由实验决定的模型称作**经验模型**(empirical model). 通贯全书, 我们将介绍把所设计实验的结果转化为所研究系统的经验模型的技巧. 这些经验模型像机理模型一样能被科学家和工程师使用.

一个设计良好的实验是重要的, 因为从实验中得到的结果和结论在很大程度上依赖于收集数据的方法. 为了说明这一点, 假定在上述实验中那位冶金工程师用来自同一炉的试件进行油淬, 而用第二炉的试件进行盐水淬. 那么, 当比较平均硬度时, 工程师就不能说出所观测到的差别有多少是淬火介质所致, 以及有多少是由于炉次的不同而致①. 因此, 收集数据的方法对由实验所能得到的结论起到了负面的影响.

一般地, 实验可以用来研究过程和系统的性能. 过程或系统可以用图 1.1 所示的模型来表示. 通常可以形象地将过程视为操作、机器、方法、人以及其他资源的一种组合, 它把某种输入 (通常是一种物质) 转变为有一个或多个可观测的响应(response) 变量的一种输出. 某些过程变量和材料特性 $x_1, x_2, \cdots, x_p$ 是可控制的, 而另一些变量 $z_1, z_2, \cdots, z_q$ 是不可控制的 (尽管它们对试验的目的来说也许是可控的). 实验的目的包括:

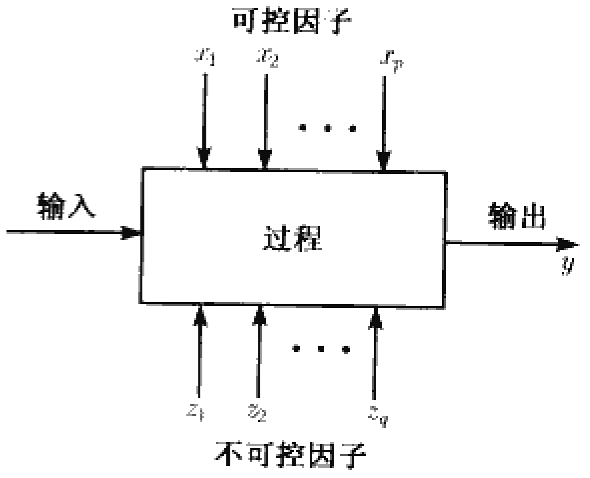


图 1.1 过程或系统的一般模型

(1) 确定哪些变量对响应 y 的影响最大;

(2) 确定有影响的 x 设置为何值可使 y 几乎接近于所希望的额定值;

(3) 确定有影响的 x 设置为何值可使得 y 的变异性较小;

(4) 确定有影响的 x 设置为何值可使得不可控制的变量 $z_1, z_2, \cdots, z_q$ 的效应最小.

正如前面讨论所见, 实验往往包括几个因子. 通常进行实验的人 (称为**实验者**) 的目的, 就是要确定这些因子对系统的输出响应的影响. 设计和进行实验的一般方法称作**实验策略**. 实验者可以采取多个策略. 我们将用一个非常简单的例子来说明一些策略.

我很喜欢打高尔夫球. 可是我不喜欢练习, 所以我总是在寻找一种降低得分的更简单的办法. 我认为有些因子可能会很重要, 即可能会影响我的高尔夫得分, 这些因子包括: (1) 使用的长打棒的类型 (特大尺寸的还是常规尺寸的); (2) 使用的球的类型 (树胶的还是三片的); (3) 步行并手提高尔夫球棒还是乘高尔夫球车; (4) 打球时喝的是水还是啤酒; (5) 在上午打球还是在下午打球; (6) 打球时天气是冷还是热; (7) 高尔夫球鞋钉的类型 (金属的还是软的); (8) 在有风的日子还是无风的日子打球.

显然, 还可以考虑其他的因子. 假定我们主要感兴趣的就是这些因子. 而且, 基于长期的比赛经验, 我认为第 5~8 个因子可以忽略, 这些因子之所以不重要是因为它们的效应非常小, 没有实际意义. 工程师和科学家通常必须对他们正考虑的真实实验中的部分因子作出以上的决策.

现在, 考虑怎样通过实验检验第 1~4 个因子对我的高尔夫成绩的影响. 假定实验过程中最

① 实验设计专家可能会说, 淬火介质和炉次的效应已经混杂了. 也就是说, 这两种因子的效应不能分开了.

多能打 8 轮高尔夫球. 一种方法就是选择这些因子的任意组合, 测试它们, 观测发生了什么. 例如, 选择使用特大长打棒、树胶球、乘高尔夫球车和喝水的组合, 结果为 87 杆. 但是在这一轮中, 我注意到使用特大长打棒有几杆进球结果并不好 (高尔夫比赛中长并不总是好的), 因此我决定在另一轮中使用常规长打棒, 其他因子保持不变. 基于目前试验的结果, 下次试验改变一个 (或两个) 因子的水平, 这种方法可以几乎无限期地继续下去. 这种我们称之为**最佳猜测法**(best-guess approach) 的实验策略在实践中经常被工程师和科学家所采用. 因为实验者具有大量关于他们正研究的系统的技术上或理论上的知识, 还有相当丰富的实际经验, 所以该方法的使用效果通常相当好. 但是最佳猜测法至少有两点不足. 第一, 假定最初的最佳猜测没有产生预期的结果, 实验者不得不做另一种关于因子水平正确组合的猜想, 这要继续很长一段时间, 而且没有成功的保证. 第二, 假定最初的最佳猜测产生了一个可以接受的结果, 现在虽然不能保证最优解决办法已被发现, 但实验者仍试图终止实验.

另一种在实际中被广泛应用的实验策略是**一次一因子**(one-factor-at-a-time) 方法. 这种方法包括对每个因子选择初始点, 或者水平的**初始**(baseline) 组合, 然后在保持其他因子在初始水平不变的条件下, 让每一个因子在其所允许的范围内进行连续变动. 当所有的试验都做完后, 我们就可以作出一系列的图形, 来显示响应变量如何受各个单因子变化 (即保持其他因子不变) 的影响. 图 1.2 给出了高尔夫实验的一系列图, 将特大长打棒、树胶球、步行和喝水作为 4 个因子的初始水平. 这个图的解释是非常直观的, 例如, 行进模式曲线的斜率是负的, 我们可以得出乘车有助于提高成绩的结论. 根据这些一次一因子图形, 我们选择的最优组合是常规长打棒、乘车和喝水. 高尔夫球的类型似乎并不重要.

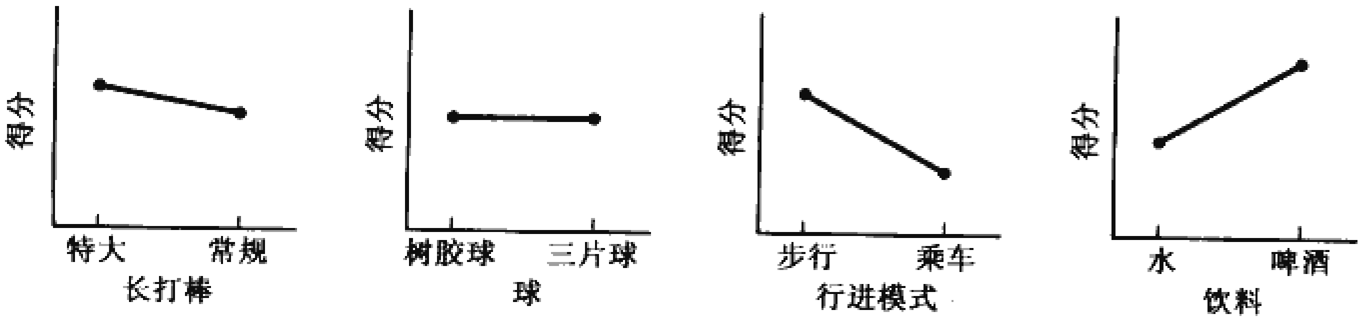


图 1.2 高尔夫实验的一次一因子策略的结果

一次一因子策略的主要缺点在于, 它没有考虑因子间可能存在的**交互作用**(interaction). 交互作用会使一个因子与另一个因子的不同水平的结合使用难以对响应产生同样的效应. 图 1.3 说明了高尔夫实验中长打棒的类型和饮料因子之间的交互作用. 可以看出, 如果我使用常规长打棒, 所喝的饮料类型对得分几乎没有影响, 但是如果我使用特大长打棒, 喝水的效果比喝啤酒的效果更好. 因子间的交互作用是非常普遍的, 如果交互作用存在, 那么一次一因子的策略产生的结果往往不理想. 许多人没有意识到这一点, 结果在实际中经常采用一次一因子实验.(有些人确实认为这个策略是一种科学方法, 或者认为它"合乎"工程原理.) 然而对设

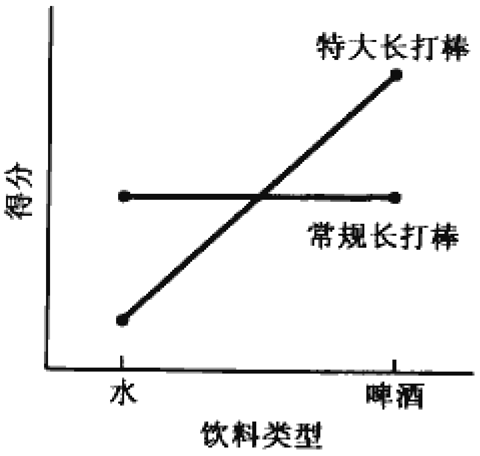


图 1.3 高尔夫实验长打棒类型和球的类型的交互作用

计而言, 一次一因子实验往往比其他基于统计途径的方法效率更低. 第 5 章将详细讨论这一点.

处理多个因子的正确方法是进行**析因实验**(factorial experiment). 这种实验策略是所有因子一起变化, 而不是一次变一个. 析因实验设计的概念非常重要, 因此本书中有好几章都介绍基本析因实验及其大量有用的变形和特例.

为了说明析因实验是如何进行的, 考虑高尔夫实验, 假定只考虑两个因子: 长打棒的类型和球的类型. 图 1.4 显示了研究两个因子对高尔夫得分的联合效应的二因子析因实验. 注意到这个析因实验的两个因子都有两个水平, 而且两个因子水平交叉的所有可能组合在设计中都用上了. 从几何上看, 也就是 4 轮实验的结果对应于一个正方形的 4 个角点. 这种特殊的析因实验称作$\mathbf{2^2}$**析因设计**(两个因子, 每个因子有两个水平). 因为我希望打 8 轮高尔夫球来检测这些因子, 所以图 1.4 中显示的每个因子水平组合可以各打 2 轮. 实验设计者称之为**重复**设计两次. 这样的实验设计可以帮助实验者研究每个因子的个体效应 (或主效应) 并判定因子是否存在交互作用.

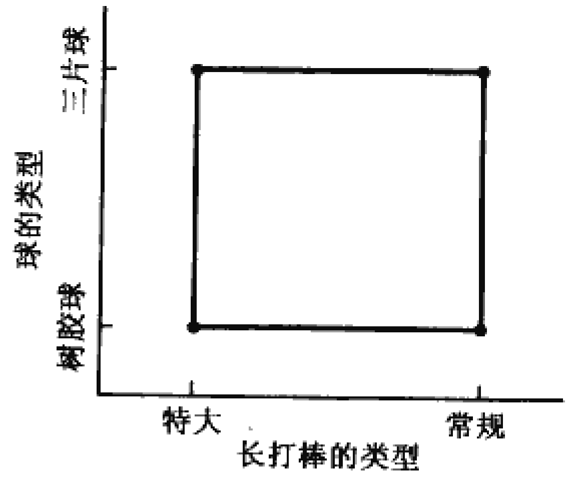


图 1.4 涉及长打棒类型和球类型的二因子析因实验

图 1.5a 显示了执行图 1.4 中的析因实验的结果. 在正方形的 4 个角点上显示了每一轮高尔夫得分. 注意到有 4 个轮次的高尔夫得分提供了使用常规长打棒的信息, 另外 4 轮提供了使用特大长打棒的信息. 通过计算正方形右边和左边的得分均值差 (如图 1.5b 所示), 我们得到从特大长打棒换到常规长打棒效应的度量, 即

$$\text{球棒效应} = \frac{92+94+93+91}{4} - \frac{88+91+88+90}{4} = 3.25$$

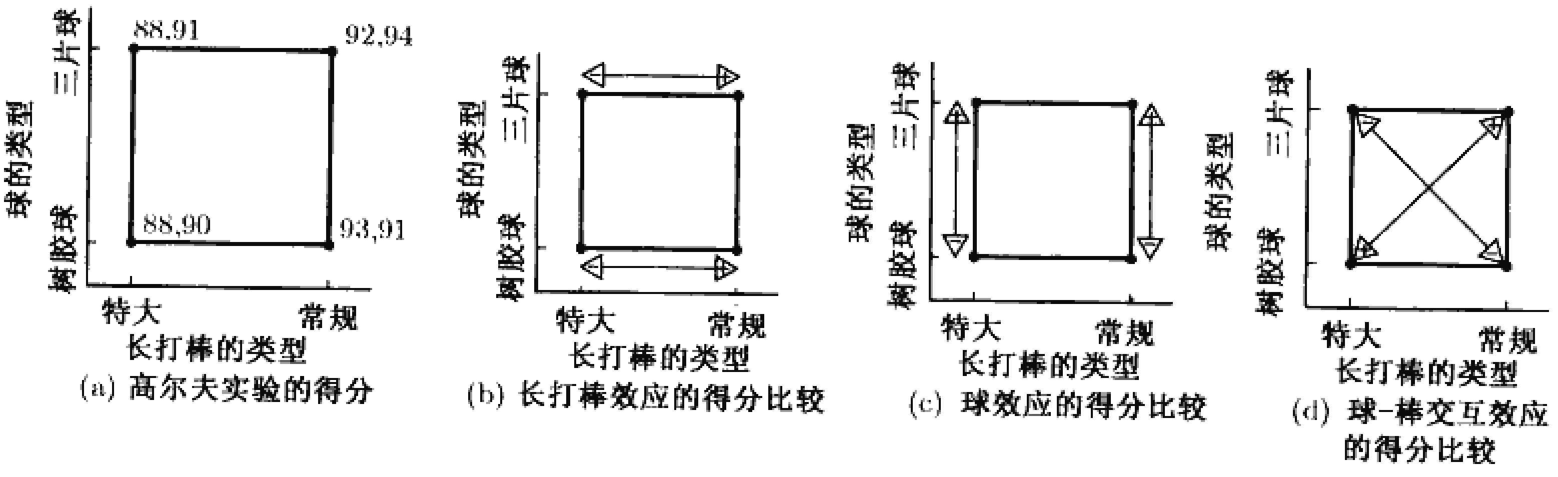


图 1.5 图 1.4 中的高尔夫实验的得分以及因子效应的计算

因此, 从特大长打棒换到常规长打棒平均每轮增加 3.25 分. 与此类似, 正方形顶部的 4 个得分和正方形底部的 4 个得分的均值差度量了所用球的效应 (见图 1.5c):

$$\text{球的效应} = \frac{88+91+92+94}{4} - \frac{88+90+93+91}{4} = 0.75$$

最后, 用右上–左下对角线和左上–右下对角线的均值差来度量球类型和长打棒类型的交互作用

(见图 1.5d), 结果如下:

$$\text{球和球棒的交互作用} = \frac{92+94+88+90}{4} - \frac{88+91+93+91}{4} = 0.25$$

析因实验的结果表明长打棒的效应大于球以及交互作用的效应. 可以利用统计检验判定这些效应与 0 是否有显著差异. 事实上, 统计证据表明长打棒效应与 0 有显著差异, 而其他两个效应与 0 没有显著差异. 所以, 实验结果提示我应该采用特大长打棒.

这个简单的例子显现出析因实验的一个非常重要的特性, 即析因实验可以高效率地利用实验数据. 注意到这组实验包括 8 个观测值, 所有观测值都用于计算长打棒、球和交互作用的效应. 没有其他的实验策略可以这样高效率地利用数据, 这是析因实验一个重要且有用的特性.

可以将析因实验的概念推广到 3 个因子. 假定希望研究长打棒类型、球类型以及饮料类型对高尔夫得分的效应, 并且假定这 3 个因子都有两个水平, 析因实验的设计可以如图 1.6 所示. 注意到各有两个水平的 3 个因子有 8 个试验组合, 8 个试验点可以在几何上用立方体的 8 个顶点来表示. 这是一个**2^3析因设计**的例子. 因为我只想打 8 轮高尔夫球, 所以, 此次实验要求我每轮只能打图 1.6 中的 8 个顶点所代表的其中一种组合. 但是, 与图 1.4 中的二因子析因设计相比, 2^3 析因设计在因子的效应方面提供了相同的信息. 例如, 若图 1.4 中的二因子设计中的每轮实验重复两次, 则在这两个设计中, 都有 4 个试验提供关于常规长打棒的信息, 4 个试验提供关于特大长打棒的信息.

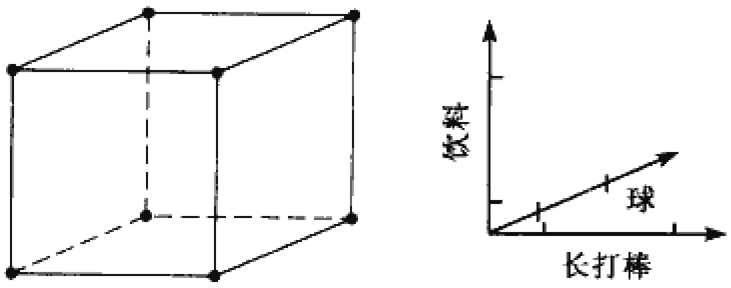


图 1.6 涉及长打棒类型、球类型以及饮料类型的三因子析因实验

图 1.7 显示了如何在**2^4析因设计**中研究所有的 4 个因子 —— 长打棒、球、饮料和行进方式 (步行或乘车). 像在任意的析因设计中一样, 它利用了因子水平组合的所有可能. 由于所有 4 个因子都有两个水平, 所以这个实验设计在几何上仍然可以用立方体 (实际上是超立方体) 表示.

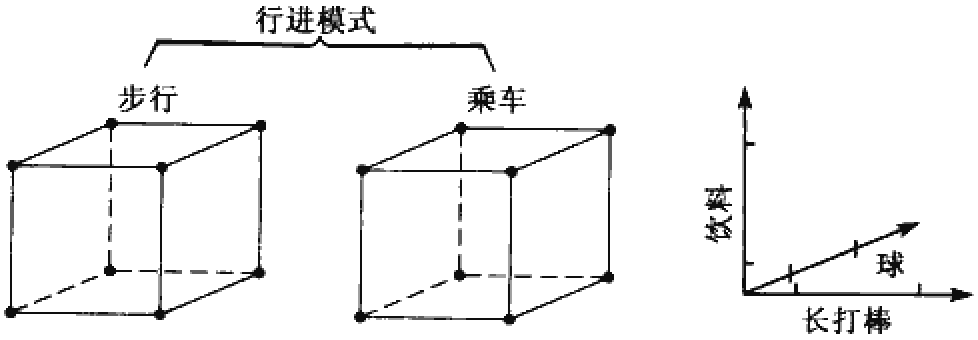


图 1.7 涉及长打棒类型、球类型、饮料类型以及行进模式的四因子析因实验

一般地, 如果有 k 个因子, 每个因子有两个水平, 那么析因设计就要进行 2^k 轮. 例如, 图 1.7 中的实验要进行 16 轮. 显然, 随着感兴趣的因子数的增加, 要求的轮数也迅速增加. 比如, 具有 10 个因子、每个因子都有两个水平的实验需要进行 1 024 轮. 这无论从时间还是资源角度来说,

都是不可行的. 在高尔夫实验中, 我只能打 8 轮高尔夫球, 所以图 1.7 中的实验次数就已经很多了.

所幸的是, 如果有 4~5 个甚至更多的因子, 通常没必要对因子水平所有可能的组合一一进行试验. **分式析因实验**(fractional factorial experiment) 是基本的析因设计的变形, 它只对所有组合的一个子集进行试验. 图 1.8 显示了高尔夫实验中 4 个因子的一个分式析因设计. 这个实验只需要做 8 轮而不是原来要求的 16 轮, 因此称作 **$\frac{1}{2}$ 分式**(one-half fraction). 如果我只做 8 轮实验, 那么这是研究所有 4 个因子的极好试验. 它提供了很好的关于 4 个因子主效应的信息以及因子间如何交互的部分信息.

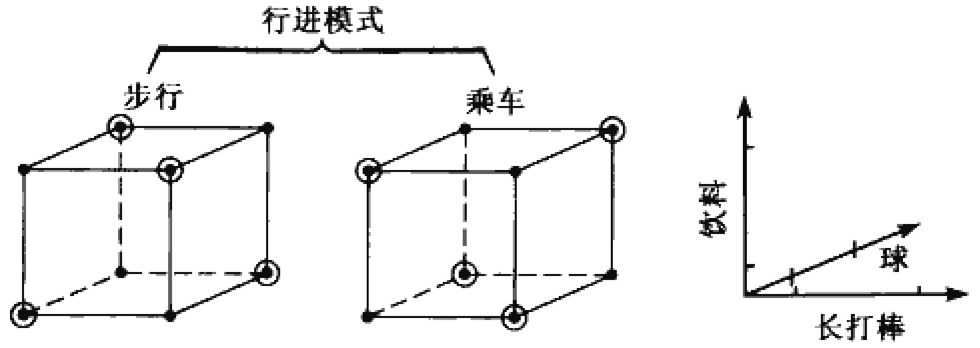


图 1.8 涉及长打棒类型、球类型、饮料类型以及行进模式的四因子分式析因实验

分式析因设计广泛应用在产品研究和开发以及过程改进中. 这些设计将在第 8 章中讨论.

1.2 实验设计的一些典型应用

实验设计方法在很多学科中得到了广泛的应用. 正如前面提到的一样, 我们可以把实验看作科学研究过程的一部分, 也可以把它看作是探究系统或过程运行方式的一种途径. 一般来说, 我们先是对某一过程提出一些猜想, 然后进行实验, 产生这一过程的一些数据, 再用来自这一实验的信息建立新的猜测, 它又导出新的实验, 如此等等. 我们正是通过这样的一系列活动进行探究的.

在科学和工程学领域, 实验设计是改进产品实现过程的非常重要的手段. 这些活动是新的制造过程设计与开发以及过程管理中的重要元素. 在过程开发的早期, 应用实验设计方法能够: (1) 提高产量; (2) 减少变异性, 与额定值或目标值更为一致; (3) 缩短开发时间; (4) 降低总成本.

实验设计方法在开发新产品、改进老产品的**工程设计**(engineering design) 活动中也扮演着重要角色. 实验设计在工程设计中的某些应用包括: (1) 评价和比较基本的设计结构; (2) 评定材料的选择方案; (3) 选择设计参数, 使得产品能在广泛的实际使用条件下运行良好, 也就是说, 使得产品是**稳健的**; (4) 确定影响产品性能的关键产品设计参数; (5) 新产品的配方. 在这些领域中, 应用实验设计可以使产品更容易制造, 具有更高的现场性能和可靠性, 具有更低的产品成本、更短的产品设计和开发时间. 现在介绍几个例子来予以说明.

例 1.1 过程刻画

一台浸流焊接机用在印刷电路板的制造过程中. 机器将电路板在焊剂中清洗、预热, 然后将它顺着传送带通过一波融化的焊料. 这一焊接过程将板上的加铅元件进行电连接和机械连接.

目前, 该过程产生大约 1%的次品. 也就是说, 电路板上约有 1%的焊点是不合格的, 需要人工修正. 但是, 因为每个印刷电路板平均含有 2 000 多条焊接线, 即使是 1%的次品水平也使需要返工的焊点太多了.

对此，负责该过程的工程师想用实验设计来确定哪些机器参数对出现焊料次品有影响，以及应该怎样调整那些变量以便减少焊料次品.

浸流焊接机有以下几个可以控制的变量：(1) 焊料温度；(2) 预热温度；(3) 传送带速度；(4) 焊剂类型；(5) 焊剂的比重；(6) 焊料波的深度；(7) 传送带角度.

除了这些可控制的因子之外，还有另外几个因子在日常生产中不容易控制 (尽管为了试验的目的可以对它们进行控制). 这些因子是：(1) 印刷电路板的厚度；(2) 电路板上所用的元件的类型；(3) 电路板上的元件的布局；(4) 操作工；(5) 生产率.

在这种情况下，工程师想要刻画 (characterize) 浸流焊接机. 也就是说，他想要确定哪些因子 (可控制的与不可控制的) 对印刷电路板出现次品有影响. 要完成这一任务，他可以设计一个实验以便能够估计因子效应的大小和方向. 即当每个因子变化时，响应变量 (每单位的次品数) 能有多大的变化，所有因子一起改变产生的结果与调节个别因子产生的结果有没有差别 (即，因子间是否存在交互作用)？有时候，我们称这类实验为**筛选实验**(screening experiment). 筛选实验或刻画实验一般需要使用分式析因设计，如图 1.8 所示的高尔夫例子.

由这种筛选实验或刻画实验所得到的信息，将用于识别关键的过程因子，并确定这些因子的调整方向以便进一步减少单位次品数. 实验也可以提供关于在日常生产中哪些因子应该更仔细地控制的信息，以避免出现高次品水平和不稳定的过程性能. 这样一来，实验的一种结果可能就是，不仅可以将控制图技术运用到过程输出上，而且可以将像控制图这样的技术应用到一个或多个**过程变量**(例如焊料温度) 中. 过一段时间，如果过程有足够的改善，就有可能主要依靠过程控制方案来控制过程输入变量，而不是依靠控制图来控制过程输出.

例 1.2 过程优化

在刻画实验中，人们通常希望确定哪些过程变量会影响过程的响应. 逻辑上，下一步骤就是优化，即确定导致最佳响应的重要因子的范围. 例如，如果响应是产率，则寻求最大产率的区域；如果响应是产品的一个重要尺度的变异性，则寻求最小变异性的区域.

假设我们想要提高某一化学过程的产率. 由刻画实验的结果知道，影响产率的两个最重要的过程变量是操作温度和反应时间. 当前过程是在温度 145°F，反应时间 2.1 小时进行的，产率约为 80%. 图 1.9 显示了时间–温度区域的俯视图. 在此图中，具有相同产率的连线形成了响应**等高线**(contour)，该图显示了产率为 60%，70%，80%，90%，95%的等高线图. 这些等高线是产率曲面 (对应于前面所说的百分比产率) 的截面在时间–温度区域上的投影. 此产率曲面有时叫做**响应曲面**(response surface). 图 1.9 中的真实响应曲面对生产人员来说是未知的，所以需要实验的方法并通过调整时间和温度来优化产率.

为了找出最优值的位置，需要进行时间和温度一起变化的实验，即，析因实验. 图 1.9 显示了时间和温度都各有两个水平的析因实验的结果. 在正方形 4 个角点处观测得到的响应显示，我们应该按增加温度且减小反应时间的方向移动以增加产率. 依此方向再进行少量的附加实验，而这些附加实验将足以使我们找到最大产率区域的位置.

一旦找到了优化区域，一般就要进行另一次实验. 第 2 次实验的目标是开发一个过程的经验模型，并得到一个对时间和温度的最优运行条件的精确估计. 这种过程最优化的方法称作**响应曲面方法**(response surface methodology)，第 11 章将会详细研究它. 图 1.9 中的第 2 个实验是一个**中心复合设计**(central composite design)，是用在过程优化研究中最重要的实验设计之一.

例 1.3 产品设计I

生化工程师要设计一种新的药品静脉推射泵，这种推射泵应该在特定的时间里注射规定剂量的药品. 她必须指定一些变量或设计参数，其中包括空筒的直径和长度、空筒和活塞的适合度、活塞的长度、连接推射泵和插入病人静脉的针的乳头的直径和壁厚、用来制造空筒和乳头的材料，以及系统必须控制的额定气压. 部分设计参数的效应可以通过建立原型 (原型中的参数在合适的范围内变化) 来评估. 可以设计实验来检验

原型, 从而发现哪些设计参数对静推泵的性能影响最大. 分析这些信息可以帮助工程师找到一种设计以提供可靠且平稳的药品注射.

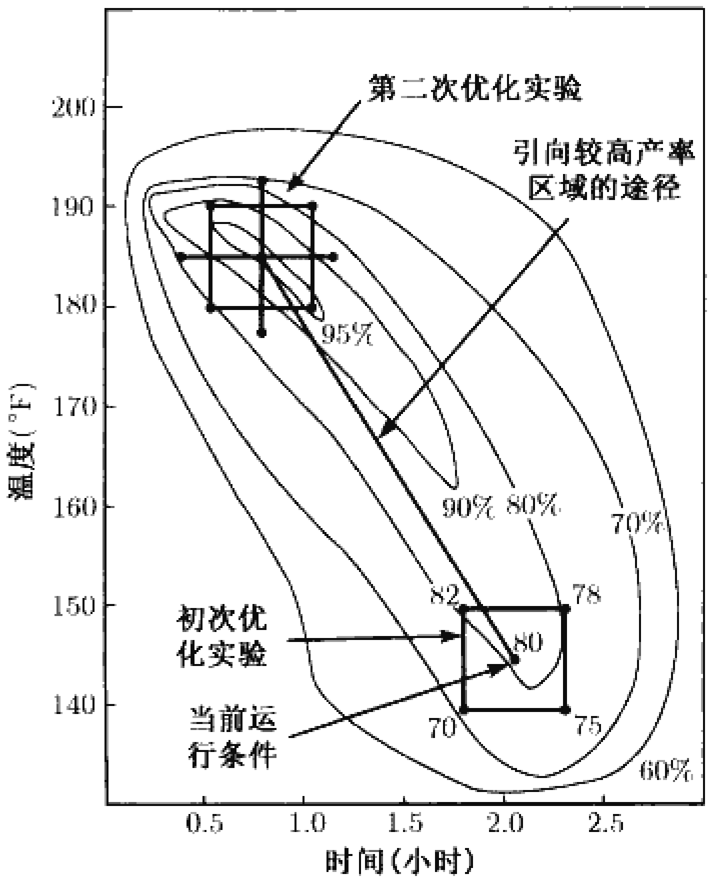


图 1.9 说明优化实验的产率作为反应时间和反应温度的函数的等高线图

例 1.4 产品设计II

工程师要设计一种飞机发动机 (商用的涡扇发动机), 希望能运用到飞行高度为 40 000 英尺、飞行速度为 0.8 马赫的巡航机中. 设计参数包括进口流量、扇压比、总气压、定子出口温度以及许多其他因子. 系统的输出响应变量有耗油量和发动机推进力. 在设计过程的早期, 不允许建造原型或试验飞机, 所以工程师利用系统的**计算机模型**(computer model), 并把重点放在发动机的关键设计参数上, 他们改变这些参数以便优化发动机的性能. 实验设计可以借助发动机的计算机模型来确定最重要的设计参数和它们的最优设置.

设计师经常利用计算机模型来辅助设计. 例如, 许多结构性设计和机理性设计方面的有限元模型, 集成电路设计的电路模拟, 工厂或企业级的用于进度计划和产量计划或供应链管理的模型, 以及复杂化学过程的计算机模型. 统计设计实验可以像在实际的物理系统中一样简便和成功地应用在这些模型当中, 从而缩减开发时间, 以获得更好的设计.

例 1.5 产品配方

生化工程师要设计一种诊断产品来检测某一种特定疾病是否存在. 该产品是生物材料、化学试剂和其他材料的混合物, 当与人的血液混合时发生反应并提供诊断信息. 这里使用的实验类型是**混料实验**(mixture experiment), 因为组成诊断产品的各成分在混合物中的比率总和为 100%(无论是体积比、重量比还是摩尔比). 响应是在产品中出现的混合比的函数. 混料实验是响应曲面实验的特殊类型, 第 11 章将会研究它们. 在设计生物科技产品、药品、油漆和涂料、消费品 (比如清洁剂、肥皂和其他个人护理用品), 以及各种其他产品方面, 它们都有很大的用途.

1.3 基 本 原 理

如果想要更有效地进行例 1.1 至例 1.5 所描述的那种实验, 就必须用科学的方法来设计它

们. 所谓**实验的统计设计**(statistical design of experiments), 就是设计实验的过程, 以便收集适合于用统计方法分析的数据, 从而得出有效且客观的结论. 如果想从数据中得出有意义的结论, 那么用统计方法作实验设计是必要的. 当问题涉及受实验误差影响的数据时, 只有统计方法才是**客观的**(objective) 分析方法. 因此, 任何实验问题就存在两个方面: 实验的设计和数据的统计分析. 这两个方面是紧密相连的, 因为分析方法直接依赖于所用的设计. 这两个论题在本书中都有涉及.

实验设计的 3 个基本原理是**随机化**(randomization)、**重复**(replication) 和**区组化**(blocking). 随机化是实验设计中使用统计方法的基石. 它的意思是, 实验材料的分配和实验中各次试验进行的顺序都是随机确定的. 统计方法要求观测值 (或误差) 是独立分布的随机变量. 随机化通常能使这一假定有效. 把实验进行适当的随机化亦有助于 "平均掉" 可能出现的外来因子的效应. 例如, 假设上述硬度实验中的试件的厚度稍有差异, 而淬火介质的效应可能会受试件厚度的影响. 如果所有油淬的试件都比盐水淬的试件厚, 则实验结论可能存在系统偏差. 这个偏差将对其中一种淬火介质不利, 从而对我们的结论也不利. 然而给淬火介质随机地分配试件就会缓解这种问题.

计算机软件程序经常用于辅助实验者选择和指导实验设计. 这些程序经常使实验设计按照随机的顺序执行每一轮. 随机顺序可以由一个随机数字发生器产生. 但即使有了这样一种计算机程序, 仍有必要安排在实验中使用的各种实验材料 (比如前面提到的硬度例子中的各个试件)、操作者和测量装置等.

有时候实验者很难随机化实验的某一部分. 例如在化学过程中, 温度是非常难以改变的变量. 相对于其他因子, 我们很少改变它. 在这种实验中, **完全随机化**是非常困难的, 因为它需增加时间和费用. 我们将采用统计设计方法来处理随机化的限制问题. 这些方法的一部分将在后续章节中讨论 (特别地, 见第 13 章).

所谓**重复**, 是指每个因子水平组合的独立重复. 在 1.1 节讨论的冶炼实验中, 重复包含用油淬处理试件和用盐水淬处理试件. 这样, 当在每种淬火介质中处理了 5 块试件时, 我们就说做了 5 次**重复**. 这 10 个观测值中的每一个都是按照随机次序进行的. 重复有两条重要的性质. 第一, 它允许实验者得到一个实验误差估计, 这个误差估计成为一个确定数据之间的观测差是否统计意义的实际差的基本度量单位. 第二, 如果用样本均值 ($\overline{y}$) 估计实验中某一因子水平的响应均值的真值, 则重复能够使得实验者得到更精确的参数估计. 例如: 如果 σ^2 是单个观测值的方差, 且有 n 次重复, 则样本均值的方差是

$$\sigma_{\overline{y}}^2 = \frac{\sigma^2}{n}$$

也就是说, 如果有 $n = 1$ 次重复并观测得 $y_1 = 145$(油淬) 和 $y_2 = 147$(盐水淬), 我们也许不能作出关于淬火介质效应的满意推断, 即观测差可能是实验误差的结果. 关键是, 没有重复我们就没有办法知道这两个观测值为何不同. 另一方面, 如果 n 合理地大且实验误差足够小, 则当我们观测到 $\overline{y}_1 < \overline{y}_2$ 时, 就有相当可靠的结论: 对这种特殊的铝合金而言, 盐水淬会比油淬产生更高的硬度.

通常在随机化实验中连续两次 (或更多次) 试验的一些因子会出现完全相同的水平. 例如, 一个实验中有 3 个因子: 压强、温度和时间. 随机化实验次序后, 我们发现:

试验号	压强 (psi)	温度 (℃)	时间 (分钟)
i	30	100	30
$i+1$	30	125	45
$i+2$	40	125	45

第 i 次和第 $i+1$ 次试验中, 压强的水平是相同的. 而第 $i+1$ 次和第 $i+2$ 次试验中, 温度和时间的水平是相同的. 为了获得真正的重复性, 实验者需要在第 i 次和第 $i+1$ 次试验中间, "扭动压力旋钮", 调到一个中间值, 并在第 $i+1$ 次试验中重新设定压强为 30 psi. 类似地, 在第 $i+1$ 次和第 $i+2$ 次中间, 在设定第 $i+2$ 次试验设计水平之前, 先将温度和时间设定在中间水平. 部分实验误差就源自达到和维持因子水平的变异性.

重复和**重复测量**是有很大区别的. 例如, 假定在一个单晶片等离子体蚀刻工序中蚀刻一个硅晶片, 晶片的关键尺寸需测量 3 次. 这种测量不是重复, 而是重复测量. 此时, 3 次重复测量的观测差异受测量系统或量具内在的变异性的直接影响. 又如, 在半导体制造实验中, 用一个氧化炉在特定的气流率和时间的条件下同时加工 4 个晶片, 测量每个晶片的氧化物厚度. 和前例相同, 这 4 个晶片的测量不是重复而是重复测量. 在这个案例中, 这 4 个测量值反映了特定的炉次中晶片间差异和变异性的其他来源. 而重复既反映试验**间**的变异又反映试验**内**(内在) 的变异.

区组化是用来提高实验精度的一种设计技术, 使用它我们在感兴趣的因子之间进行比较. 区组化常用于减少或消除**讨厌因子**带来的变异. 讨厌因子是指可能影响实验响应而我们不直接感兴趣的因子. 例如, 在一个化学过程中, 要在所有的轮次中使用两批原材料. 然而, 由于供货商之间的不同可能造成批次间的差异, 如果我们对这种影响不是特别感兴趣, 就可以将原材料的批次作为讨厌因子. 一般地, 区组是一组相对类似的实验条件. 在化学过程的例子中, 批次内的变异预期小于批次间的变异, 所以各个批次的原材料将组成一个区组. 一般情况下, 正如在该例中一样, 讨厌因子的每个水平形成一个区组. 实验者在统计设计中将观测值按区组进行分组. 本书多处 (包括第 4, 5, 7, 8, 9, 11, 13 章) 将详细地讨论区组化问题, 2.5.1 节给出了一个简单例子来说明区组化原理.

实验设计的 3 个基本原理 (随机化、重复和区组化) 是每个实验的重要部分. 通贯本书, 我们将反复地解释并强调这些基本原理.

1.4 设计实验指南

为了在设计和分析实验时使用统计方法, 与实验有关的每一个人需要预先对所研究的问题究竟是什么以及如何收集数据等有一个清晰的认识, 至少要对如何分析这些数据有一个定性的了解. 推荐的步骤大纲见表 1.1. 现在我们简要地讨论该大纲并详细地描述其中的一些要点. 更多的细节, 见 Coleman and Montgomery(1993). 本章的**补充材料**也是有用的.

(1) **问题的识别与表述** 这一点看起来似乎是再明白不过了, 但是在实践中, 确认需要实验的问题的存在却并不是那么简单的, 将问题阐述得简明而又可以被普遍接受就更不简单了. 需要对实验目的有一个全面的考虑. 通常, 吸收所有有关各方的参与是很重要的, 其中包括: 工程部、质量保证部、制造部、市场营销部、管理部门、顾客以及操作工 (通常, 他们有很多很好的想法, 却常常被忽略了). 基于这个原因, 采用**团队方法**来设计实验是值得推荐的.

表 1.1 设计实验指南

(1) 问题的识别与表述 (2) 响应变量的选择* (3) 因子、水平和范围的选择*	实验前的计划
(4) 实验设计的选择	
(5) 进行实验	
(6) 数据的统计分析	
(7) 结论和建议	

* 在实践中, 第 2 步和第 3 步通常同时进行或以相反次序进行.

对实验涉及的有关问题列一个清单往往是有帮助的. 明确地陈述问题通常有助于更好地理解正在研究的现象以及问题的最终解决方案. 牢记总目标也很重要. 例如, 它是否一个新的过程或系统 (此时初始目标可能是**刻画或因子筛选**)? 它是否一个已经被提前刻画的成熟的或者容易理解的系统 (此时目标可能是优化)? 一个实验可能有许多目标, 包括**确认**(系统运行的方式是否与以前一样?)、**发现**(如果我们开发新的材料、变量、运行条件等, 将会发生什么?)、**稳定性**或**稳健性**(在什么条件下响应变量受因子的影响小, 或者我们怎样才能减小我们不能直接控制的来源引起的响应变量的变异?). 显然, 实验中涉及的问题直接与总目标相关. 提出问题时必须认识到一个大的综合性实验不可能满意地回答所有的关键性问题. 一个综合性实验要求实验者知道许多问题的答案, 如果他们错了, 结果将是令人失望的. 这不仅浪费时间、材料和其他资源, 也许还会导致根本不能满意地回答起初研究的问题. 采用**序贯**的方法是一个较好的策略. 所谓序贯的方法, 就是做一系列较小的实验, 每个实验都有一个特定的目标, 比如因子筛选.

(2) **响应变量的选择** 在选择响应变量时, 实验者应该确信这个变量确实会对所研究的过程提供有用的信息. 我们经常取测量特性的平均值或标准差 (或两者) 为响应变量. 多重响应并非不常用. 仪表性能 (或测量误差) 也是一个重要因子. 如果仪表性能差, 则只有相对大的因子效应才能通过实验检测出来, 或者需要做附加的重复实验. 在一些仪表性能不好的情形下, 实验者可以多次测量每个实验单元, 采用重复测量的平均值作为响应的观测值. 在进行实验前, 确定与响应的定义及测量方式有关的事项是非常重要的. 有时设计实验被用来研究和改进测量系统的性能, 相关例子见第 12 章.

(3) **因子、水平和范围的选择** (如表 1.1 的注释所示, 第 2 步和第 3 步通常同时进行, 或者以相反次序进行.) 当所考虑的因子可能影响过程或系统的改进时, 实验者发现这些因子通常可以分为**潜在设计因子**或讨厌因子. 潜在设计因子是指实验者希望在实验中改变的因子. 通常我们发现会有许多的潜在设计因子, 所以对它们进行进一步分类是有帮助的. 一些有用的分类包括**设计因子**、**保持常量因子**和**允许改变因子**. 设计因子是指在实验中被实际选择用来研究的因子. 保持常量因子是指可能对响应施加一些影响的变量, 但是对于当前实验的目的而言, 我们对这些因子并不感兴趣, 所以它们将被保持在一个特定的水平. 例如, 在半导体工厂的蚀刻实验中, 也许有一个效应对于用在实验中的特定的等离子蚀刻机来说是唯一的. 但是, 这个因子在实验中是很难改变的, 所以实验者也许会在一个特定的 (在理想情况下, 是 "典型的") 蚀刻机上进行所有的实验, 因此该因子保持为常量. 至于允许改变因子, 虽然实验单元或设计因子所运用到的 "材料" 通常不是类似的, 但是我们经常忽略掉单元和单元之间的变异, 而依靠随机化来抵消材料或

实验单元的效应. 我们通常假定保持常量因子和允许改变因子的效应相对较小.

但是, 讨厌因子可能有较大的效应, 尽管在当前实验的环境中我们对此并不感兴趣, 但是必须考虑这种效应. 讨厌因子通常分为**可控因子**、**不可控因子**和**噪声因子**. 可控的讨厌因子是指它的水平可被实验者设置. 例如, 实验者在进行实验时可以选择原材料的不同批次或者一周中的不同日子. 前面讨论的区组概念经常用来处理可控的讨厌因子. 如果一个讨厌因子在实验中是不可控的, 但是它是可以测量的, 那么可以采用一个被称为**协方差分析**的分析过程来补偿它的效应. 例如, 过程环境的相对湿度影响过程的性能, 如果湿度不能控制, 那么它有可能被测量且被视为协变量. 若因子在过程中自然变化且不可控制, 但基于实验的目的而言它是可控的, 则称此因子为噪声因子. 在这种情况下, 我们的目标通常是寻找可控实验因子的设置, 以最小化噪声因子引起的变异性. 有时称之为过程稳健性研究或稳健设计问题. 区组、协方差分析和过程稳健性研究将在后面章节继续讨论.

一旦实验者选择了设计因子, 他或她必须选择这些因子变化的范围及其特定水平, 还必须考虑如何将这些因子控制在所希望的数值上以及如何测量这些数值. 例如, 在浸流焊接实验中, 工程师已经确定了 12 个可能影响出现焊接次品的变量. 实验人员还必须确定每个变量的范围 (即每个因子未来可能变化的范围) 以及每个变量使用多少个水平. 要做到这一点, 就需要一定的**过程知识**. 这种过程知识通常是实践经验和理论理解的结合. 能够研究所有的重要因子而不受过去经验的过分影响 (特别是在实验的早期阶段或过程远未成熟时) 是很重要的.

当实验目的是**因子筛选**或**过程刻画时**, 通常应选择较少的因子水平数. 一般地, 在因子筛选研究中两个水平是较好的. 选择感兴趣的范围也是重要的. 在因子筛选中, 范围必须相对地大, 即因子变化的范围应该较宽. 当我们对哪些变量重要和哪些水平能产生最佳效果认识得更多的时候, 感兴趣的范围通常就会变得窄一些.

为了组织在实验前的计划中产生的一些信息, 可以使用**因果图**(cause-and-effect diagram). 图 1.10 是一个因果图, 用于计划一个实验去解决半导体工厂的蚀刻机中遇到的晶片负荷 (电荷聚集在晶片上) 问题. 因果图也称为**鱼刺图**, 因为感光趣的 "效应" 或响应变量画在图的脊骨上, 潜在原因或设计因子安排在一串肋骨上. 因果图采用传统意义的原因去组织信息和潜在的设计因子, 包括测量、材料、人员、环境、方法和机器. 我们注意到一些单个原因可以直接作为包含在实验里的设计因子 (比如车轮转速、气流和真空), 而其他列出的潜在区域则需进一步研究, 才能将它们转为设计因子 (比如操作者不正确的工艺流程), 剩下的原因可能导致其他的因子, 它们在实验期间保持常量或者被区组化 (比如温度和相对湿度). 图 1.11 是一个实验的因果图, 用于研究数控机床上的涡轮刀片的几个因子的效应. 这个实验有 3 个响应变量: 刀剖面、刀面抛光和已抛光刀片的表面抛光缺陷. 原因可以分成几组: 可以从中选择实验设计因子的可控因子、效应可能被随机化抵消的不可控因子、可能区组化的讨厌因子和进行实验时可能保持常量的因子. 实验者在实验前的计划期间构建几个不同的因果图来帮助和指导自己是很正常的. 关于数控机床实验的更多信息, 以及在实验前计划中用到的图示方法的进一步讨论, 见本章补充材料.

我们重申, 前面从第 1 步到第 3 步所有的观点和过程信息是非常重要的. 我们把这归诸于**实验前计划**. Coleman and Montgomery(1993) 一书中提供了实验前计划的很有用的工作表. 更多的细节和使用这些工作表的一个例子也见本书的**补充材料**. 在许多情况下, 一个人不可能拥有所需的全部知识来把这类事情做得充分好. 因此, 我们强烈呼吁在计划实验过程中的团队效应.

成功的关键大多取决于实验前计划的好坏.

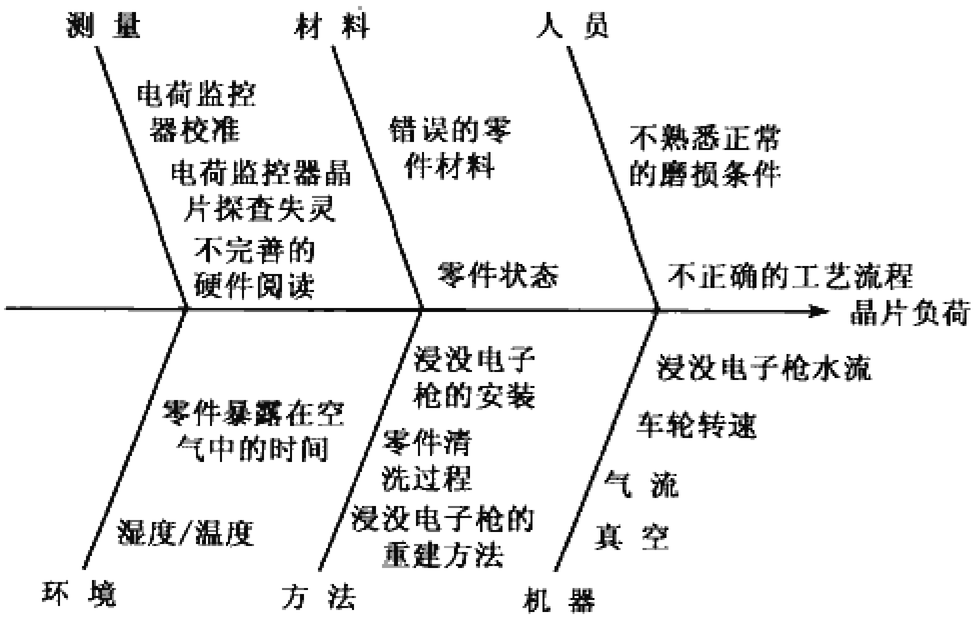


图 1.10 蚀刻工序实验的因果图

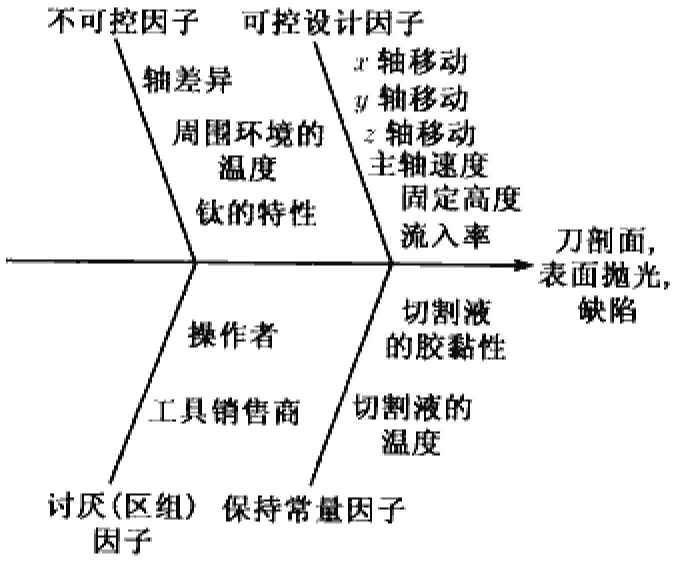


图 1.11 数控机床实验的因果图

(4) **实验设计的选择** 如果正确地做了前面的实验前计划活动, 这一步就相对容易些. 设计的选择涉及考虑样本量 (重复次数), 选择合适的实验次序, 确定是否划分区组或者是否涉及其他随机化限制. 本书讨论一些较重要的实验设计类型, 对于大多数问题来说, 你都可以从中选择到合适的实验设计类型.

在实验设计的这个阶段, 有许多交互式的统计软件包支持. 实验者可以输入包括因子数、水平和范围的信息, 这些程序将会给出一个可供考虑的设计选择或推荐一个特定的设计. (在大多数情况下, 我们更愿意看到一些选择方案而不是依赖计算机的推荐.) 这些程序通常也会提供一个工作表 (带已随机化的试验次序) 用于指导实验的进行.

在选择设计时, 时刻记着实验的目的是很重要的. 在很多工程实验中, 我们已经知道, 有些因子水平会使响应得出不同的数值. 因此, 我们想要识别是哪些因子引起差异并估计响应改变量的大小. 在其他情况下, 我们会更注意验证一致性. 例如, 比较两种生产条件 A 和 B, A 是标准的, B 是成本效益更好的. 实验者可能会想弄清楚这两种生产条件的产率是否没有差异.

(5) **进行实验** 进行实验时, 需要仔细监控实验的过程以确保每件事情都按计划做完. 在这

个阶段, 实验程序中的错误通常会破坏实验的有效性. 而预先计划是成功的关键. 在复杂的制造或研究开发环境中实施已设计好的实验, 人们很容易低估物流方面和计划方面的问题.

Coleman and Montgomery(1993) 认为进行实验前, 做少数试验或尝试性试验通常是有帮助的. 这些试验提供了关于实验材料一致性的信息, 使你能够检查测量系统, 对实验误差有一个粗略估计, 以及有机会实践全部的实验技术. 如果必要, 可以借此机会重新审视在第 1~4 步中所作的决定.

(6) **数据的统计分析** 分析数据应该用统计方法, 以便结果和结论都是客观的, 而不是主观臆断的. 如果实验设计正确, 并且按设计执行了, 则所需的统计方法不必去费心准备. 因为有很多优秀的软件包可用来分析数据, 许多用在第 4 步中去选择设计的程序提供了无缝的、直接的界面进行统计分析. 在数据的分析和解释中, 简单的**图解法**起着重要的作用. 因为实验者想回答的许多问题可以纳入假设检验的框架, 所以假设检验和置信区间估计程序非常有助于分析已设计好了的实验的数据. 它通常也有助于根据经验模型来给出许多实验的结果. 经验模型是指从数据中推导出的等式, 它表达了响应和重要设计因子之间的关系. 残差分析和模型适合性检测也是重要的分析方法. 后面的章节将会详细地讨论这一点.

统计方法并不能证明一个因子 (或几个因子) 有特殊的效应. 它们仅对实验结果的可靠性和有效性提供准则. 只要统计方法应当得当, 虽然我们不能用实验来证明任何事情, 但是却可以用它们度量结论中可能出现的误差或者对一个命题附加上置信水平. 统计方法的主要优点是它对决策过程加进了客观性. 统计方法与良好的工程知识 (或过程知识) 以及常识相结合通常有助于人们得出正确的结论.

(7) **结论和建议** 分析过数据之后, 实验者必须作出有关实验结果的真实结论, 并推荐一种处理方法. 这时候, 图解法常常比较有用, 特别是在给别人介绍成果时更是如此. 另外, 还需要进行**跟踪试验**与**确认试验**以证实实验结论的正确性.

通贯整个过程, 要牢记实验是学习过程的一个重要部分, 在学习过程中, 我们暂时提出了关于系统的假设, 进行实验来研究这些假设, 根据实验的结果再提出新的假设, 如此等等. 这表明, 实验是**迭代式**地逐步深化的. 通常一种错误的做法, 是在研究一开始, 就去设计一个单一、庞大和内容广泛的实验. 一个成功的实验需要先弄清其中的重要因子, 这些因子可能变化的整个范围, 使用合适的水平个数, 以及度量这些变量的合适单位. 一般说来, 我们不可能完全知道这些问题的答案, 但是, 当我们不断实验下去就会获得对于它们的更多认识. 随着实验的进展, 我们经常会抛弃一些输入变量, 而加进一些其他变量, 改变某些因子的研究范围, 或者加进新的响应变量. 因此, 我们通常**序贯地**进行试验. 作为一般法则, 在第一次试验中, 投入的可用资源不要超过约 25%. 这样可以确保有足够的资源用来进行确认试验并最终完成实验的最后目的.

1.5 统计设计简史

现代统计实验设计的发展分为 4 个时期. 在 20 世纪 20 年代和 30 年代初期, 开拓者 Ronald A. Fisher 爵士引领了农业时期. 在那个时期, Fisher 在英国伦敦附近的 Rothamsted 农业实验站负责统计和数据分析. 他认识到, 生成数据的实验过程的缺陷通常会影响系统 (这里是指农业系统) 的数据分析. 通过与许多领域的科学家和研究者交流讨论, 他总结出 1.3 节里讨论过的 3

个实验设计的基本原理：随机化、重复和区组化. Fisher 系统地将统计的思想和原理引入到实验设计研究, 包括析因设计概念和方差分析. 他的两本书 [Fisher(1958, 1966)] 对统计的应用产生了深刻影响, 特别在农业领域以及相关的生命科学领域. 关于 Fisher 的一部优秀的传记, 见 Box(1978).

尽管统计设计在工业背景下的应用确实开始于 20 世纪 30 年代, 但是第二时期或者工业时期是由 Box and Wilson(1951) 所提出的响应曲面方法 (RSM) 催生的. 他们认识到许多工业实验与对应的农业实验有两点根本不同：(1) 响应变量通常可以被 (几乎) 立即观测到, (2) 实验者可以很快从一小组实验中得到用于规划下个实验的重要信息. Box(1999) 称工业实验的这两个特性为**及时性**和**序贯性**. 在接下来的 30 年中, RSM 和其他设计技术在整个化工邻域和处理行业中被广泛应用, 且大多数用于研究和开发工作. George Box 明智地领导了这项运动. 然而, 统计设计在工厂或制造过程生产上的应用还不十分广泛. 部分原因是对于工程师或其他过程专家在基本的统计概念和方法上的培训不够, 而且缺少计算资源和用户友好的统计软件的支持.

20 世纪 70 年代后期, 西方工业对质量改进兴趣的增加开启了统计设计的第三时期. 田口玄一的著作 [Taguchi and Wu (1980), Kackar(1985), Taguchi(1987, 1991)] 使人们对于实验设计的兴趣和应用的推广具有显著性的影响. 田口主张在实验设计中使用他提出的**稳健参数设计**, 即

(1) 使过程对环境因子或难以控制的其他因子不敏感;

(2) 使产品对元部件的变异不敏感;

(3) 寻找过程变量的水平使均值达到目标值, 同时减少围绕该值的变异.

田口特别推荐使用分式析因设计和其他正交排列以及一些新的统计方法去解决这些问题. 结果, 在方法论上引起了更多的争论. 引起争论的部分原因是, 虽然田口的方法在西方首先 (或者主要) 被企业家提倡, 但根本的统计科学没有得到西方同行足够的审评. 到 20 世纪 80 年代末, 同行的评价表明, 尽管田口的工程理念和目标是完善的, 但是他的实验策略和数据分析的方法有大量的问题. 这些论点的细节见 Box(1988), Box, Bisgaard, and Fung(1988), Hunter(1985,1989), Myers and Montgomery (2002), 以及 Pignatiello and Ramberg(1992). 涉及的大多数内容也被概括总结在 *Technometrics* 杂志 1992 年 5 月的广泛小组讨论中 [见 Nair 等人 (1992)].

有关田口的争论有许多肯定的结果. 首先, 实验设计可以更广泛地用在零部件工业, 包括汽车、航空航天制造、电子和半导体以及其他一些以前很少采用这种技术的产业. 第二, 标志着统计设计的第四时期的开始. 在这个时期, 研究者和实践者都在统计设计上有了全新的、综合的兴趣, 工业领域开发了许多新的、有用的解决实验问题的方法, 其中包括田口技术方法的替代方法, 使他的工程理念可以更经济、更有效地应用于实践. 这些替代方法将在后面的章节中讨论和说明, 特别是在第 12 章. 第三, 统计实验设计的正规教育已成为许多大学里大学生和研究生工程课程的一部分. 良好的实验设计经验与工程学和科学的成功结合已成为未来工业竞争的关键因子.

实验设计的运用已远不止农业领域. 没有一个科学和工程领域不能成功地使用实验的统计设计. 最近几年, 许多其他领域也在考虑利用实验设计, 包括商业服务部门、财务服务、政府运作, 以及一些非营利事务部门. 1996 年 3 月 11 日的《财富 (福布斯)》杂志上发表了题为 “The New Mantra:MVT” 的文章, 这里 MVT 代表的是 “多元变量检验”, 作者用它来描述析因设计. 文章提及许多不同类型的公司通过运用统计实验设计取得成功的范例.

1.6 小结: 在实验中使用统计方法

在工程学、科学和工业中, 大多数的研究工作都是以经验为根据的, 并且广泛使用实验方法. 统计方法能够大大提高这些实验的效率, 并经常强化得出的结论. 在实验中恰当地使用统计方法就要求实验者牢记下列几点.

(1) **利用你对问题非统计知识的了解**. 实验者通常精通于他们自己的专业. 例如, 专门研究水文学问题的工程师在这一领域内一般都有着丰富的实践经验和正规的科学训练. 在某些领域, 有一大堆的物理学理论用来论述因子和响应之间的关系. 这类非统计学方面的知识, 在选择因子、确定因子水平、 决定进行多少次重复、解释分析结果等方面是极有价值的. 仅仅使用实验设计是无法替代这里应该考虑的问题的.

(2) **使设计和分析尽可能地简单**. 不要过分热心于使用复杂的、 过于精致的统计方法. 相对简单的设计和分析方法往往是最好的. 此处我们重新强调 1.4 节中所推荐的步骤 1~3. 如果你实验前计划做得很仔细并且选择了正确的设计, 那么分析方法相对来说往往是直截了当的. 事实上, 一个设计完美的实验有时几乎是自我分析! 但是, 如果你做出的实验前计划很马虎, 并且执行的实验设计又很差, 那么即使是用最复杂、最精巧的统计方法, 也几乎不可救药.

(3) **识别实际的显著性和统计的显著性之间的差别**. 仅仅因为两种实验条件产生统计上不同的平均响应, 还不能保证这一差别大得足以有实际价值. 例如, 一位工程师可以确定一种汽车燃料喷射系统的修正方法能够确实改进汽车汽油行驶的里程数, 平均为 0.1 英里/加仑, 这是一个在统计学上有意义的结果. 但是, 如果修正方法耗费 1 000 美元, 则 0.1 英里/加仑的差别可能就太小了, 没有任何实际价值.

(4) **实验通常是迭代的**. 在大多数情况下, 在研究的开始阶段, 设计内容太广泛的实验是不明智的. 成功的设计需要你了解重要的因子、 这些因子变化的范围、每个因子合适的水平数, 以及对每个因子和响应而言合适的组合方法和度量单位. 一般说来, 我们在实验之初并不能完满地回答这些问题, 但是在实验进程中, 我们可以不断深化对问题的认识. 这表明了前面讨论过的迭代方法或序贯方法的优越性. 当然, 也存在完全适用的内容广泛的实验的情况, 但是, 作为一般原则, 大多数的实验都应该是迭代的. 因此, 在初始设计中, 投入的实验资源 (试验数、预算、时间等) 通常不应该大于 25%. 第一次的努力只是在于取得经验, 而另外一些资源必须用于完成实验的最终目标.

1.7 思 考 题

1.1 假定你想设计一个实验来研究爆米花的未爆谷粒比例. 完成 1.4 节中实验设计指南的步骤 1~3. 有无难以控制的主要变异来源?

1.2 假定你想研究影响煮米饭的潜在的因子:

(a) 在这个实验中响应变量是什么? 怎样测量这个响应?

(b) 列出可能影响响应的潜在变异来源.

(c) 完成 1.4 节中的实验设计指南的前面 3 个步骤.

1.3 假定你想比较花在不同条件 (阳光、水、肥料以及土壤条件) 下的生长情况. 完成 1.4 节中的实验设计指南的步骤 1~3.

1.4 选择一个你感兴趣的实验. 完成 1.4 节中的实验设计指南的步骤 1~3.

第 2 章　简单比较实验

本章纲要

2.1　引言
2.2　基本统计概念
2.3　抽样与抽样分布
2.4　关于均值差的推断, 随机化设计
　2.4.1　假设检验
　2.4.2　样本量的选取
　2.4.3　置信区间
　2.4.4　$\sigma_1^2 \neq \sigma_2^2$ 的情形
　2.4.5　σ_1^2 与 σ_2^2 为已知的情形
　2.4.6　均值与已知值的比较
　2.4.7　小结
2.5　关于均值差的推断, 配对比较设计
　2.5.1　配对比较问题
　2.5.2　配对比较设计的优点
2.6　正态分布方差的推断
第 2 章补充材料
S2.1　数据模型和 t 检验
S2.2　模型参数的估计
S2.3　t 检验的回归模型
S2.4　构造正态概率图
S2.5　更多关于 t 检验的假设检验
S2.6　一些关于配对 t 检验的更多信息

本章将比较两种**条件**(有时称为**处理**) 的实验. 这些实验通常称作**简单比较实验**(simple comparative experiment). 我们从这样一个实验的例子开始, 通过它来确定两种不同的产品配方是否会得出相同的结果. 下面的讨论要求我们先回顾一些基本的统计概念, 例如随机变量、概率分布、随机样本、抽样分布, 以及假设检验等.

2.1　引　　言

工程师要研究硅酸盐水泥砂浆的配方. 他在混合时加入聚合物乳胶液, 判定它是否影响固化时间和砂浆的粘结抗拉强度. 实验者分别收集了原始配方的 10 个观测值和改良配方的 10 个观测值. 我们称这两种不同的配方为两种**处理**或者配方**因子**的两个**水平**. 当固化过程完成时, 实验者的确发现改良配方的砂浆的固化时间大大减少. 这时, 他开始关注砂浆的粘结抗拉强度. 如果新的砂浆配方在砂浆的粘结抗拉强度方面有反作用, 那么这将影响它的使用.

粘结抗拉强度数据列于表 2.1, 而图 2.1 用点来表示这些数据. 这种图称为**点图**. 仔细地观测这些数据, 我们的直接印象是, 未改良砂浆的强度要大于改良砂浆的强度. 改良砂浆的平均粘结抗拉强度为 $\overline{y}_1 = 16.76\ \text{kgf/cm}^2$, 而未改良砂浆的平均粘结抗拉强度为 $\overline{y}_2 = 17.04\ \text{kgf/cm}^2$, 这两个数据也证实了我们的印象是对的. 这两个样本的平均粘结抗拉强度之间的差似乎是个不大不小的量. 然而, 这并不表明这个差大得足以证明这两个配方确实是有差异的. 也许观测得到的平均强度的差别是由于抽样的波动而产生的, 而实际上这两个配方的粘结抗拉强度却是相同的. 如果另外给出两个样本也许会得出相反的结论, 即改良砂浆的强度大于未改良配方的强度.

一种称之为**假设检验**(hypothesis testing, 有时称**显著性检验**, significance testing) 的统计推断方法可用来帮助实验者比较这两种配方. 假设检验允许在客观的条件下来比较两种配方, 但是存在着得出错误结论的风险. 在提出简单比较实验中所用的假设检验方法之前, 我们简要概括

一些基本的统计概念.

表 2.1 硅酸盐水泥配方的粘结抗拉强度数据

j	改良砂浆 y_{1j}	未改良砂浆 y_{2j}	j	改良砂浆 y_{1j}	未改良砂浆 y_{2j}
1	16.85	16.62	6	17.04	16.87
2	16.40	16.75	7	16.96	17.34
3	17.21	17.37	8	17.15	17.02
4	16.35	17.12	9	16.59	17.08
5	16.52	16.98	10	16.57	17.27

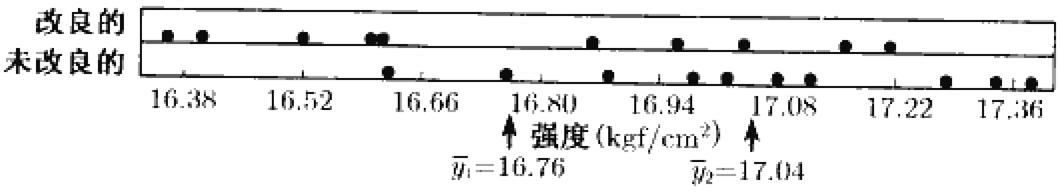


图 2.1 表 2.1 中粘结抗拉强度数据的点图

2.2 基本统计概念

上述硅酸盐水泥实验中的每一次观测称为一次**试验**(run). 每次试验都是有区别的, 所以在试验结果上存在波动 (又叫作**噪音**). 这种噪音通常称为**实验误差**或称为**误差**(error). 这是一种**统计误差**, 是由于不可控制的变异而引起的, 因而通常是不可避免的. 误差或噪音的出现意味着相应的变量 (即粘结抗拉强度) 是一个**随机变量**(random variable). 随机变量可以是**离散型**(discrete) 的, 也可以是**连续型**(continuous) 的. 如果随机变量所可能取值的集合是有限的或可数无限的, 则此随机变量是离散型的; 如果随机变量所可能取值的集合是一个区间, 则此随机变量是连续型的.

1. 变异性的图示法

我们常用简单的图示法来辅助分析由实验所得的数据. 如图 2.1 所示的**点图**(dot diagram), 是用来显示一组小样本数据 (比如说, 大约 20 个观测值) 的一种很有用的方法. 这种点图能使实验者快速地看出观测值的总体**位置**或**中心趋势**(central tendency) 及其**分散程度**(spread). 例如, 在硅酸盐水泥粘结抗拉实验中, 点图显示出两种配方在平均强度上可能有差别, 但两种配方所得的强度变异性则大体上是相同的.

如果数据比较多, 则点图中的点就会难于区分, 此时就不如采用直方图了. 图 2.2 表示 200 个观测值的直方图, 这些观测值是在一个冶炼过程中回收 (或生产) 金属所得到的. 这一直方图显示出这组数据的中心趋势、分散程度及其分布的一般形状. 直方图是这样做成的: 在水平轴上划分为若干个小区间 (通常是等长的), 在第 j 个区间上画一个矩形, 使矩形的面积与 n_j 成正比, 其中 n_j 是观测值中落入第 j 个区间中观测值的个数.

盒图 (或**带触点盒图**)是显示数据的一种很有用的方法. 它在一个以水平方向和垂直方向定位的矩形盒上显示最小值、最大值、下四分位数 (quartile) 与上四分位数 [分别代表第 25 个百分位数 (percentile) 和第 75 个百分位数], 以及中位数 (第 50 个百分位数). 盒子是从下四分位

数伸展至上四分位数, 在盒内用一条直线段表示中位数. 一般情况下, 从盒的两端各引一条直线段 (或带触点的线段) 分别至最小值与最大值. [许多盒图对样本端点的表示有不同的规定, 更多细节见 Montgomery and Runger(2003)].

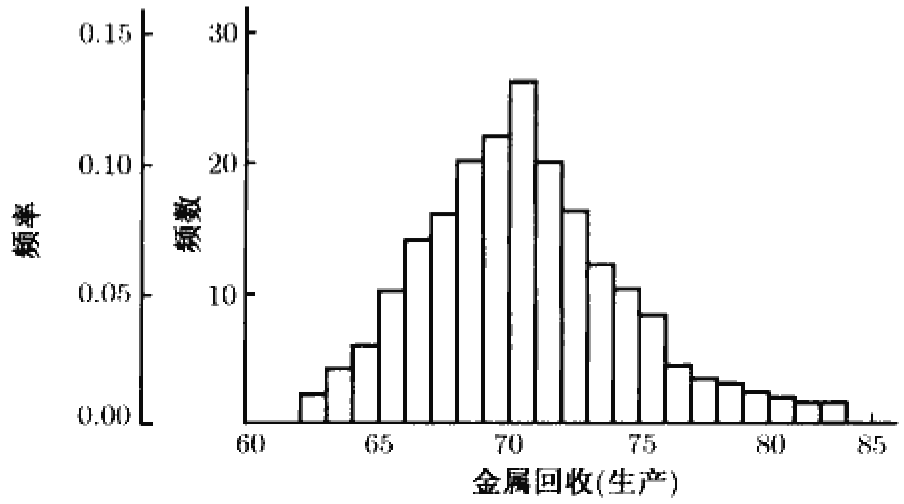


图 2.2 冶炼过程中回收 (生产) 金属所得到的 200 个观测值的直方图

图 2.3 显示了硅酸盐水泥砂浆实验中粘结抗拉强度的两个样本的盒图. 该图表明两种配方之间平均强度的差别. 它也表明两种配方的强度分布相当对称, 其变异性或分散程度也很相似.

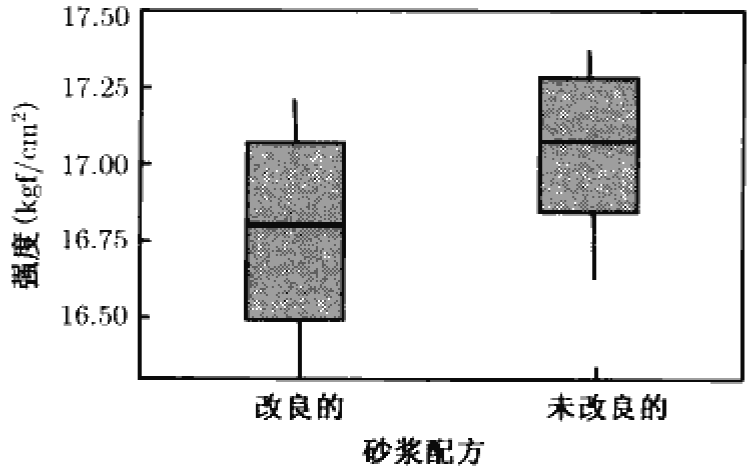


图 2.3 硅酸盐水泥砂浆粘结抗拉强度实验的盒图

点图、直方图以及盒图都可用来概括出一个数据样本(sample) 的信息. 要更全面地描述可能出现在样本中的观测值, 则要用到概率分布的概念.

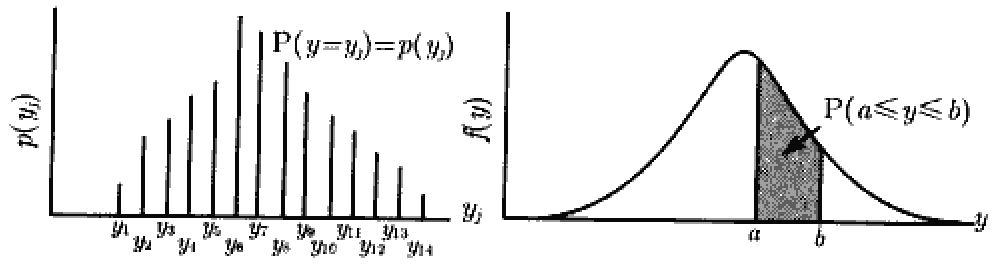
2. 概率分布

一个随机变量 y 的概率结构是用它的**概率分布**(probability distribution) 来描述的. 如果 y 是离散型的, 我们常用 y 的概率函数 [记为 $p(y)$] 来刻画 y 的概率分布. 如果 y 是连续型的, 则常用 y 的概率密度函数 [记为 $f(y)$] 来刻画 y 的概率分布.

图 2.4 描述了假想的离散型和连续型的概率分布. 注意在离散型概率分布中, 它是用概率函数 $p(y)$ 的高度来表示相应的概率; 而在连续型概率分布中, 它是用曲线 $f(y)$ 下相应于给定区间上的面积来表示落入该区间内的概率. 概率分布的性质可以定量地概述如下.

$$y\text{为离散型: 对于一切}y_j\text{的值,}\quad 0 \leqslant p(y_j) \leqslant 1,\quad \mathrm{P}(y = y_j) = p(y_j),\quad \sum p(y_j) = 1.$$

$$y\text{为连续型: } f(y) \geqslant 0,\quad \mathrm{P}(a \leqslant y \leqslant b) = \int_a^b f(y)\mathrm{d}y,\quad \int_{-\infty}^{\infty} f(y)\mathrm{d}y = 1.$$



(a) 离散型分布 (b) 连续型分布

图 2.4 离散型与连续型概率分布

3. 均值、方差和期望值

一个概率分布的**均值**(mean)μ 是它的中心趋势或中心位置的度量. 数学上, 均值定义为

$$\mu = \begin{cases} \int_{-\infty}^{\infty} yf(y)\mathrm{d}y, & y\text{为连续型} \\ \sum\limits_{\text{一切}y} yp(y), & y\text{为离散型} \end{cases} \tag{2.1}$$

也可以用随机变量 y 的**期望值**(expected value) 或随机变量 y 的长期试验的平均值来表示均值, 即

$$\mu = \mathrm{E}(y) = \begin{cases} \int_{-\infty}^{\infty} yf(y)\mathrm{d}y, & y\text{为连续型} \\ \sum\limits_{\text{一切}y} yp(y), & y\text{为离散型} \end{cases} \tag{2.2}$$

其中 E 表示**期望值算子**.

一个概率分布的分散程度或者离中趋势, 可以用**方差**(variance) 来度量, 方差定义为

$$\sigma^2 = \begin{cases} \int_{-\infty}^{\infty} (y-\mu)^2 f(y)\mathrm{d}y, & y\text{为连续型} \\ \sum\limits_{\text{一切}y} (y-\mu)^2 p(y), & y\text{为离散型} \end{cases} \tag{2.3}$$

方差可以由期望来表示, 因为有关系式

$$\sigma^2 = \mathrm{E}[(y-\mu)^2] \tag{2.4}$$

最后, 因为方差被广泛应用, 为了方便, 我们定义一个**方差算子**V:

$$\mathrm{V}(y) = \mathrm{E}[(y-\mu)^2] = \sigma^2 \tag{2.5}$$

期望值与方差的概念广泛应用于本书, 因此, 温习一下关于这些算子的几个基本结果也许对我们会有帮助. 如果 y 是具有均值 μ 和方差 σ^2 的随机变量, c 是一个常量, 则

(1) $\mathrm{E(c)} = \mathrm{c}$

(2) $\mathrm{E}(y) = \mu$

(3) $\mathrm{E(c}y) = \mathrm{cE}(y) = \mathrm{c}\mu$

(4) $\mathrm{V(c)} = 0$

(5) $\mathrm{V}(y) = \sigma^2$

(6) $V(cy) = c^2V(y) = c^2\sigma^2$

如果两个随机变量, 例如 y_1 和 y_2, 分别有 $E(y_1) = \mu_1$ 和 $V(y_1)=\sigma_1^2$, 以及 $E(y_2)=\mu_2$ 和 $V(y_2)=\sigma_2^2$, 则有

(7) $E(y_1 + y_2)=E(y_1)+E(y_2)=\mu_1+\mu_2$

可以证明

(8) $V(y_1 + y_2)=V(y_1)+V(y_2)+2Cov(y_1, y_2)$

其中

$$Cov(y_1, y_2) = E[(y_1 - \mu_1)(y_2 - \mu_2)] \tag{2.6}$$

是随机变量 y_1 与 y_2 的**协方差**(convariance). 协方差用来度量 y_1 与 y_2 之间的线性相关性. 特别地, 可以证明, 如果 y_1 与 y_2 独立①, 则 $Cov(y_1, y_2)=0$. 还可证明

(9) $V(y_1 - y_2)=V(y_1)+V(y_2)-2Cov(y_1, y_2)$

如果 y_1, y_2**独立**, 则有

(10) $V(y_1 \pm y_2)=V(y_1) + V(y_2) = \sigma_1^2 + \sigma_2^2$

(11) $E(y_1 \cdot y_2) = E(y_1) \cdot E(y_2) = \mu_1 \cdot \mu_2$

然而要注意, 一般说来, 无论y_1 与 y_2 是否独立,

(12) $E\left(\frac{y_1}{y_2}\right) \neq \frac{E(y_1)}{E(y_2)}$

2.3 抽样与抽样分布

1. 随机样本、样本均值和样本方差

统计推断的目标是, 我们要用总体的一个样本来得出关于该总体的一些结论. 我们将要研究的大多数方法都假定采用**随机样本**(random sample). 也就是说, 如果总体含有 N 个元素, 选取其中 n 个元素作为一个样本, 并且 $N!/[(N-n)!n!]$ 个可能的样本中的每一个都具有相等的概率被选中, 这样的抽样方法叫做**随机抽样**. 但在实际操作中, 有时难以得出随机样本, 而通过计算机程序也许有助于产生随机数.

统计推断充分地利用从样本观测值中计算得到的量. 我们定义**统计量**(statistic) 为样本观测值的不含有未知参数的函数. 例如, 设 $y_1, y_2, \cdots, y_n$ 表示一个样本, 则**样本均值**(sample mean)

$$\overline{y} = \frac{\sum\limits_{i=1}^{n} y_i}{n} \tag{2.7}$$

与**样本方差**(sample variance)

$$S^2 = \frac{\sum\limits_{i=1}^{n} (y_i - \overline{y})^2}{n-1} \tag{2.8}$$

都是统计量. 这些量分别是样本中心趋势和离中趋势的度量. $S = \sqrt{S^2}$, 称为**样本标准差**, 有时作为离中趋势的度量. 工程师们常常喜欢用标准差来量度离中趋势, 因为它的单位与变量 y 的单位相同.

① 注意, 这一点反过来并不成立, 亦即 $Cov(y_1, y_2)=0$ 并不隐含 y_1 与 y_2 的独立性. 例如, 见 Hines, Montgomery, Goldsman, and Borror(2003).

2. 样本均值和样本方差的性质

样本均值 $\bar{y}$ 是总体均值 μ 的一个点估计 (量), 样本方差 S^2 是总体方差 σ^2 的一个点估计 (量). 一般说来, 一个未知参数的**估计量**(estimator) 就是对应于那个参数的统计量. 注意点估计 (量) 是一个随机变量. 从样本数值所算得的估计量的具体数值叫做**估计值**(estimate). 例如, 假设要估计某一特殊类型纺织纤维抗断强度的均值和方差. 检验一个 $n = 25$ 的纤维试件的随机样本, 并记录好每一试件的抗断强度. 按照公式 (2.7) 与公式 (2.8) 计算出样本均值和样本方差, 分别是 $\bar{y} = 18.6$ 和 $S^2 = 1.20$. 因此, μ 的估计值是 $\bar{y} = 18.6$, σ^2 的估计值是 $S^2 = 1.20$.

一个良好的点估计需要具备许多性质, 其中两条最重要的性质如下.

(1) 点估计应该是**无偏的** (unbiased). 也就是说, 点估计的长期试验平均值或期望值应该是被估计的参数. 虽然无偏性是所希望的, 但仅有这一性质常常不能使一个估计量成为一个好的估计量.

(2) 一个无偏估计量应该具有**最小方差**. 这一性质表明, 最小方差点估计的方差比参数的任一其他估计量的方差都小.

容易证明, $\bar{y}$ 与 S^2 分别是 μ 与 σ^2 的无偏估计量.

先考虑 $\bar{y}$. 利用期望的性质, 得

$$E(\bar{y}) = E\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(y_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

因为每个观测值 y_i 的期望值都是 μ, 所以, $\bar{y}$ 是 μ 的一个无偏估计量.

现在考虑样本方差 S^2. 有

$$E(S^2) = E\left[\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right] = \frac{1}{n-1}E(SS)$$

其中 $SS = \sum_{i=1}^{n}(y_i - \bar{y})^2$ 是这些观测值 y_i 的**校正平方和**(corrected sum of square). 现在

$$\begin{aligned} E(SS) &= E\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right] \qquad (2.9)\\ &= E\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right] = \sum_{i=1}^{n}(\mu^2 + \sigma^2) - n(\mu^2 + \sigma^2/n)\\ &= (n-1)\sigma^2 \qquad (2.10)\end{aligned}$$

因此,

$$E(S^2) = \frac{1}{n-1}E(SS) = \sigma^2$$

所以 S^2 是 σ^2 的一个无偏估计.

3. 自由度

(2.10) 式中的量 $n-1$ 叫做平方和 SS 的**自由度**(degrees of freedom). 这是一个很一般的结果. 也就是说, 如果 y 是一个具有方差 σ^2 的随机变量, 而 $SS = \sum(y_i - \bar{y})^2$ 有 v 个自由度, 则

$$\mathrm{E}\left(\frac{SS}{v}\right)=\sigma^2 \tag{2.11}$$

平方和的自由度等于平方和中独立元素的个数. 例如, (2.9) 式中的 $SS=\sum_{i=1}^{n}(y_i-\overline{y})^2$ 由 n 个元素 $y_1-\overline{y}, y_2-\overline{y},\cdots,y_n-\overline{y}$ 的平方和组成, 但是 $\sum_{i=1}^{n}(y_i-\overline{y})=0$, 所以这些元素并非都是独立的. 实际上, 只有其中的 $n-1$ 个是独立的, 所以 SS 有 $n-1$ 个自由度.

4. 正态分布及其他抽样分布

如果已知样本所在总体的概率分布, 通常我们能够确定一个特定统计量的概率分布. 统计量的概率分布叫做**抽样分布**(sampling distribution). 现在, 我们简要讨论几个有用的抽样分布.

最重要的一个抽样分布就是**正态分布**(normal distribution). 如果 y 是一个正态随机变量, 则 y 的概率分布是

$$f(y)=\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-(1/2)[(y-\mu)/\sigma]^2}, -\infty<y<\infty \tag{2.12}$$

其中 $-\infty<\mu<\infty$ 是分布的均值, $\sigma^2>0$ 是方差. 正态分布如图 2.5 所示.

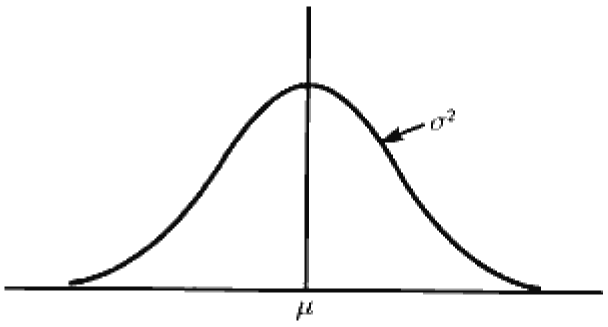


图 2.5 正态分布

因为作为实验误差的结果而体现在样本中的差别通常能用正态分布来描述, 所以正态分布在实验设计的数据分析中起着极为重要的作用. 很多重要的抽样分布亦可通过正态随机变量来定义. 我们常用记号 $y\sim N(\mu,\sigma^2)$ 表示 y 是具有均值 μ 和方差 σ^2 的正态分布.

正态分布的一个重要的特殊情况是**标准正态分布**(standard normal distribution). 即 $\mu=0$ 并且 $\sigma^2=1$. 如果 $y\sim N(\mu,\sigma^2)$, 则随机变量

$$z=\frac{y-\mu}{\sigma} \tag{2.13}$$

服从标准正态分布, 记为 $z\sim N(0,1)$. (2.13) 式所示的运算通常叫做正态随机变量 y 的**标准化**(standardizng). 累积的标准正态分布见附录的表 1.

很多统计方法都假定随机变量服从正态分布. 中心极限定理常作为近似正态性的一个合理依据.

定理 2.1 中心极限定理

若 $y_1, y_2, \ldots, y_n$ 是 n 个独立同分布的随机变量, 具有 $\mathrm{E}(y_i)=\mu$ 和 $\mathrm{V}(y_i)=\sigma^2$(都是有限的), 且 $x=y_1+y_2+\cdots+y_n$, 则当 $n\to\infty$ 时,

$$z_n=\frac{x-n\mu}{\sqrt{n\sigma^2}}$$

分布的极限形式服从标准正态分布.

这一结果说明, n 个独立同分布随机变量的和是近似正态分布的. 在很多情况中, 这一近似对很小的 n(比方说 $n < 10$) 是良好的, 但在其他情况下, n 需要很大 (比方说 $n > 100$). 我们通常认为, 实验中的误差来自多个独立源, 且以加法方式出现, 因此正态分布就成为这种联合实验误差的一个近乎合理的模型.

能利用正态随机变量定义的一个重要的抽样分布是**卡方分布**或 **χ^2 分布**. 若 $z_1, z_2, \cdots, z_k$ 是独立的正态分布的随机变量, 其均值为 0, 方差为 1, 简记为 NID(0, 1), 则随机变量

$$x = z_1^2 + z_2^2 + \cdots + z_k^2$$

服从自由度为 k 的 χ^2 分布. χ^2 的密度函数是

$$f(x) = \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} x^{(k/2)-1} \mathrm{e}^{-x/2}, \quad x > 0 \tag{2.14}$$

几个 χ^2 分布如图 2.6 所示. 这一分布是非对称的或**偏峰的**(skewed), 其均值与方差分别为

$$\mu = k, \quad \sigma^2 = 2k$$

χ^2 分布的百分位数表见附录中的表 III.

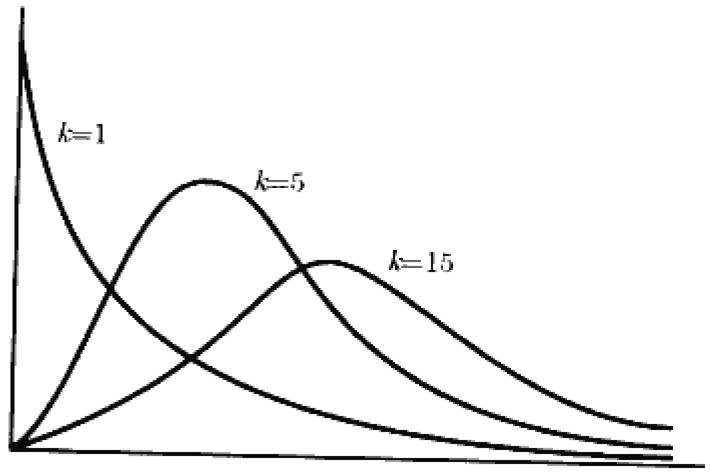


图 2.6 几个卡方分布

举一个服从卡方分布的随机变量的例子, 假设 $y_1, y_2, \ldots, y_n$ 是来自 $N(\mu, \sigma^2)$ 分布的一个随机样本, 则

$$\frac{SS}{\sigma^2} = \frac{\sum_{i=1}^{n} (y_i - \overline{y})^2}{\sigma^2} \sim \chi_{n-1}^2 \tag{2.15}$$

也就是说, $\frac{SS}{\sigma^2}$ 服从自由度为 $n-1$ 的卡方分布.

本书用到的很多方法涉及平方和的计算和变换. (2.15) 式给出的结果是十分重要的且会反复出现, 正态随机变量的平方和除以 σ^2 服从卡方分布.

由 (2.8) 式可知, 样本方差可写为

$$S^2 = \frac{SS}{n-1} \tag{2.16}$$

如果样本的观测值是 NID(μ, σ^2), 则 S^2 的分布是 $[\sigma^2/(n-1)]\chi_{n-1}^2$. 于是, 如果总体是正态分布的, 则样本方差的抽样分布是卡方分布的常数倍.

若 z 与 χ_k^2 独立, 分别是独立标准正态随机变量与卡方随机变量, 则随机变量

$$t_k = \frac{z}{\sqrt{\chi_k^2/k}} \tag{2.17}$$

服从**自由度为 k 的 t 分布**, 记为 t_k. t 的密度函数是

$$f(t) = \frac{\Gamma[(k+1)/2]}{\sqrt{k\pi}\Gamma(k/2)} \frac{1}{[(t^2/k)+1]^{(k+1)/2}}, \quad -\infty < t < \infty \tag{2.18}$$

当 $k > 2$ 时, t 的均值和方差分别是 $\mu = 0$ 与 $\sigma^2 = k/(k-2)$. 几个 t 分布如图 2.7 所示. 如 $k = \infty$, 则 t 分布变为标准正态分布. t 分布的百分位数表在附录的表 II 中给出, 若 $y_1, y_2, \cdots, y_n$ 是来自 $N(\mu, \sigma^2)$ 分布的一个随机样本, 则量

$$t = \frac{\overline{y} - \mu}{S/\sqrt{n}} \tag{2.19}$$

是自由度为 $n-1$ 的 t 分布.

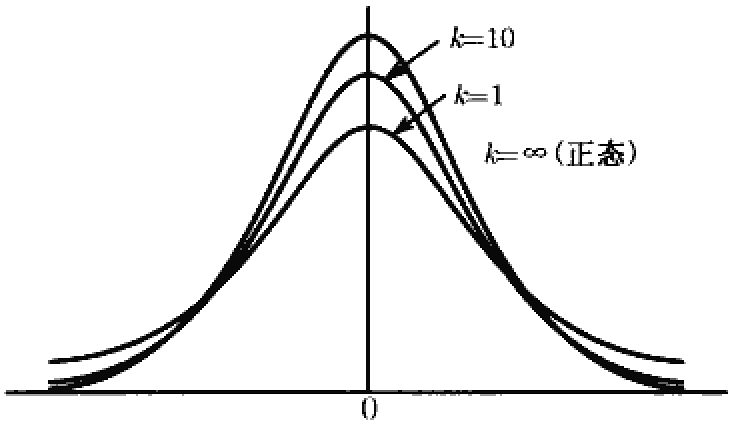


图 2.7 几个 t 分布

我们要考虑的最后一个抽样分布是 ***F* 分布**. 若 χ_u^2 与 χ_v^2 是两个独立的卡方随机变量, 其自由度分别为 u 与 v, 则比值

$$F_{u,v} = \frac{\chi_u^2/u}{\chi_v^2/v} \tag{2.20}$$

服从分子(numerator) 自由度为 u, 分母(denominator) 自由度为 v 的 F 分布. 如果 x 是分子自由度为 u, 分母自由度为 v 的 F 随机变量, 则 x 的概率分布是

$$h(x) = \frac{\Gamma\left(\frac{u+v}{2}\right)\left(\frac{u}{v}\right)^{u/2} x^{(u/2)-1}}{\Gamma\left(\frac{u}{2}\right)\Gamma\left(\frac{v}{2}\right)\left[\left(\frac{u}{v}\right)x+1\right]^{(u+v)/2}}, \quad 0 < x < \infty \tag{2.21}$$

几个 F 分布如图 2.8 所示. 这一分布在实验设计的统计分析中是十分重要的. F 分布的百分位数表见附录的表 IV.

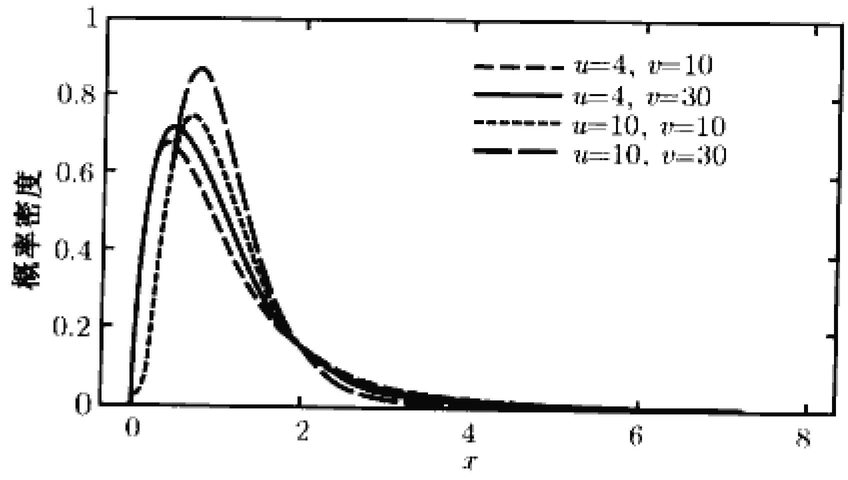


图 2.8　几个 F 分布

举一个分布为 F 的统计量的例子, 考虑两个具有共同方差 σ^2 的独立正态总体. 若 y_{11}, $y_{12}, \cdots, y_{1n_1}$ 是第 1 个总体中的 n_1 个观测值的随机样本, $y_{21}, y_{22}, \cdots, y_{2n_2}$ 是第 2 个总体中的 n_2 个观测值的随机样本, 则

$$\frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1} \tag{2.22}$$

其中 S_1^2 与 S_2^2 是两个样本方差. 这一结果直接由 (2.15) 式与 (2.20) 式得出.

2.4　关于均值差的推断, 随机化设计

现在回到 2.1 节中提出的硅酸盐水泥砂浆问题上来. 那里研究过两种不同的砂浆配方的粘结抗拉强度是否有所差别的问题. 本节将利用比较两个处理均值的**假设检验**(hypothesis testing)和**置信区间**(confidence interval) 方法来分析简单比较实验的数据.

通贯本节, 我们假定采用了一个**完全随机化实验设计**. 在这类设计中, 数据都被看作是来自正态分布的一个随机样本.

2.4.1　假设检验

现在, 我们再考虑 2.1 节中提出的硅酸盐水泥砂浆问题, 我们的兴趣在于比较两种不同配方(一个未改良砂浆的配方和一个改良砂浆的配方) 的强度, 一般地, 我们将这两种配方视为**因子**“配方”的两个**水平**. 令 $y_{11}, y_{12}, \cdots, y_{1n_1}$ 代表 n_1 个第 1 个因子水平的观测值, $y_{21}, y_{22}, \cdots, y_{2n_2}$ 代表 n_2 个第 2 个因子水平的观测值. 我们假定这是从两个独立正态总体中随机抽取的样本. 图 2.9 说明了这一点.

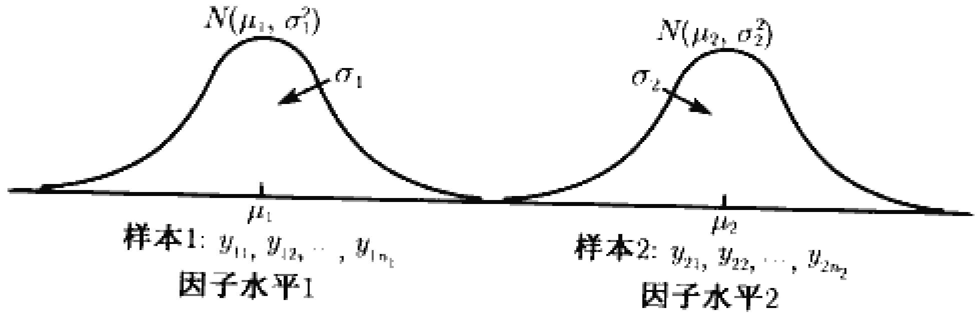


图 2.9　两样本 t 检验的抽样情形

1. 数据模型

我们通常用**模型**描述实验的结果. 描述实验中数据的简单统计模型为:

$$y_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i = 1, 2 \\ j = 1, 2, \cdots, n_i \end{cases} \tag{2.23}$$

其中 y_{ij} 是第 i 个因子水平的第 j 次观测. μ_i 是第 i 个因子水平的响应均值, ε_{ij} 是与第 ij 个观测值相关联的正态随机变量. 假定 ε_{ij} 服从 $\text{NID}(0, \sigma_i^2)$, i=1,2. 通常将 ε_{ij} 视作模型的**随机误差**(random error) 成分. 因为均值 μ_1 和 μ_2 是常量, 正如之前假定的一样, 我们可以直接从模型中看到 y_{ij} 服从 $\text{NID}(\mu_i, \sigma_i^2)$, $i = 1, 2$. 关于数据模型的更多信息, 请查阅补充材料.

2. 统计假设

统计假设(statistical hypothesis) 是一个关于概率分布参数或模型参数的命题. 这些假设反映了一些关于问题情形的**推测**(conjecture). 例如, 在硅酸盐水泥实验中, 我们可以设想, 两种砂浆配方的平均粘结抗拉强度是相等的. 这一点可以从形式上陈述为

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

其中 μ_1 是改良砂浆的平均粘结抗拉强度, μ_2 是未改良砂浆的平均粘结抗拉强度. 命题 $H_0 : \mu_1 = \mu_2$ 称为**零假设**(null hypothesis), 命题 $H_1 : \mu_1 \neq \mu_2$ 称为**备择假设** (alternative hypothesis). 此处所说的备择假设称为**双边备择假设**, 因为 $\mu_1 < \mu_2$ 时以及 $\mu_1 > \mu_2$ 时命题都为真.

为了检验假设, 我们设计一个程序来获取一个随机样本, 计算一个适当的**检验统计量**(test statistic), 然后拒绝或者不拒绝零假设 H_0. 这个程序的一部分用来找出检验统计量的一个导致拒绝 H_0 的数值集合. 该数值集合叫做检验的**临界域**(critical region) 或**拒绝域**(rejection region).

检验假设时可能会发生两类错误. 如果当零假设为真时被拒绝了, 则产生第 I 类错误. 如果零假设不真时却未被拒绝, 则出现第 II 类错误. 这两类错误的概率用特定的符号表示为:

$$\alpha = \text{P}(\text{第 I 类错误}) = \text{P}(\text{拒绝}H_0|H_0\text{为真})$$

$$\beta = \text{P}(\text{第 II 类错误}) = \text{P}(\text{未拒绝}H_0|H_0\text{为伪})$$

有时, 用检验的势来操作更为方便, 其中

$$\text{势} = 1 - \beta = \text{P}(\text{拒绝}H_0|H_0\text{为伪})$$

假设检验的一般程序是规定第 I 类错误的概率值 α, 通常叫做检验的**显著性水平**(significance level), 然后设计检验程序, 使第 II 类错误的概率 β 的值适当小.

3. 两样本 t 检验

设想我们能假设两种砂浆配方的粘结抗拉强度的方差是相同的, 则在完全随机化设计中用于比较两个处理均值的一个恰当的检验统计量是

$$t_0 = \frac{\overline{y}_1 - \overline{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{2.24}$$

其中 $\overline{y}_1$ 与 $\overline{y}_2$ 是样本均值, n_1 与 n_2 是样本量, S_p^2 是公共方差 $\sigma_1^2=\sigma_2^2=\sigma^2$ 的一个估计量, 其计算式为

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \tag{2.25}$$

S_1^2 与 S_2^2 是两个总体各自的样本方差. 为了决定是否拒绝 $H_0:\mu_1=\mu_2$, 我们将 t_0 与自由度为 n_1+n_2-2 的 t 分布进行比较. 如果 $|t_0|>t_{\alpha/2,n_1+n_2-2}$, 就拒绝$H_0$, 得出的结论是硅酸盐水泥砂浆两种配方的平均强度是不相同的, 其中 $t_{\alpha/2,n_1+n_2-2}$ 是自由度为 n_1+n_2-2 的 t 分布的上 $\alpha/2$ 百分位数. 这个检验方法通常称作**两样本 t 检验**.

这一方法的证明如下. 如果我们是从独立正态分布中进行抽样, 则 $\overline{y}_1-\overline{y}_2$ 的分布是 $N[\mu_1-\mu_2,\sigma^2(1/n_1+1/n_2)]$. 于是, 若 σ^2 已知且 $H_0:\mu_1=\mu_2$ 为真, 则

$$z_0=\frac{\overline{y}_1-\overline{y}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \tag{2.26}$$

的分布是 $N(0, 1)$. 然而, 以 S_p 代替 (2.26) 式中的 σ 时, 就将 z_0 的分布从标准正态分布转为自由度为 n_1+n_2-2 的 t 分布了. 若 H_0 为真, 则 (2.24) 式中的 t_0 分布是 $t_{n_1+n_2-2}$, 因此, 我们能期望 t_0 值的 $100(1-\alpha)\%$落在 $-t_{\alpha/2,n_1+n_2-2}$ 与 $t_{\alpha/2,n_1+n_2-2}$ 之间. 如果零假设为真, 那么一个样本得出的 t_0 值在这一范围之外时就是不正常的, 也就表明 H_0 应该被拒绝. 这样自由度为 n_1+n_2-2 的 t 分布是检验统计量 t_0 的合适的**参考分布**(reference distribution), 因而当零假设为真时, 它描述的是 t_0 的表现. 注意 α 是这一检验法的第 I 类错误的概率.

在有些问题中, 人们可能希望只要一个均值大于另一个均值时就拒绝 H_0. 于是提出了一个**单边备择假设**$H_1:\mu_1>\mu_2$, 只要 $t_0>t_{\alpha,n_1+n_2-2}$ 就拒绝 H_0. 如果人们希望只要 μ_1 小于 μ_2 就拒绝 H_0, 则备择假设是 $H_1:\mu_1<\mu_2$, 只要 $t_0<t_{\alpha,n_1+n_2-2}$ 就拒绝 H_0.

为了具体说明这一方法, 考虑表 2.1 中的硅酸盐水泥数据. 由这些数据得

改良砂浆	未改良砂浆
$\overline{y}_1=16.76\ \text{kgf/cm}^2$	$\overline{y}_2=17.04\ \text{kgf/cm}^2$
$S_1^2=0.100$	$S_2^2=0.061$
$S_1=0.316$	$S_2=0.248$
$n_1=10$	$n_2=10$

因为样本方差是相似的, 所以可以认为这两个总体的总体标准差 (或方差) 是相等的. 因而, 可以运用 (2.24) 式来检验假设

$$H_0:\mu_1=\mu_2,\qquad H_1:\mu_1\neq\mu_2$$

更进一步, $n_1+n_2-2=10+10-2=18$, 如果选择 $\alpha=0.05$, 那么当检验统计量的值 $t_0>t_{0.025,18}=2.101$ 或 $t_0<-t_{0.025,18}=-2.101$ 时, 我们就拒绝 $H_0:\mu_1=\mu_2$. 这些临界区域的边界表示在图 2.10 的参考分布 (自由度为 18 的 t 分布) 上.

由 (2.25) 式, 有

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}=\frac{9(0.100)+9(0.061)}{10+10-2}=0.081$$

$$S_p = 0.284$$

检验统计量是

$$t_0 = \frac{\overline{y}_1 - \overline{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{16.76 - 17.04}{0.284\sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{-0.28}{0.127} = -2.20$$

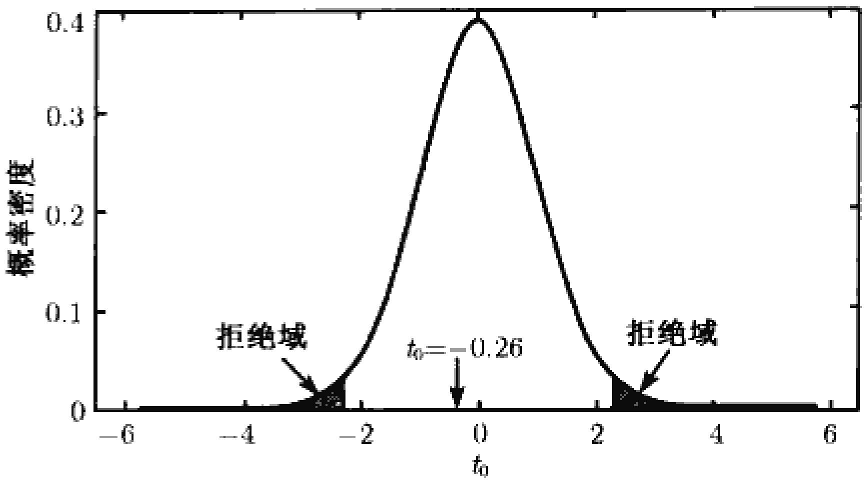


图 2.10 自由度为 18 的 t 分布的拒绝域 $\pm t_{0.025,18} = \pm 2.101$

因为 $t_0 = -2.20 < -t_{0.025,18} = -2.101$, 所以我们拒绝 H_0 并且得出硅酸盐水泥砂浆的两种配方的平均粘结抗拉强度是不同的结论. 这是一个潜在的重要工程发现. 砂浆配方的改良已经收到了缩短固化时间的效果, 而又有证据表明这个改变也影响了粘结抗拉强度. 于是, 可以得出结论: 改良配方减小了粘结抗拉强度 (正因为我们做的是双边检验, 当拒绝零假设时, 我们不能预先推出单边的结论). 如果粘结抗拉强度均值的减少在实际中也很重要 (或者除统计显著性外, 还有工程显著性), 那么还需要做更多的研究工作和进一步的实验.

4. 假设检验中的 P 值方法

报告假设检验结论的一种方式是, 在给定 α 值或者说**显著性水平**下拒绝还是不拒绝零假设. 例如, 在前面的硅酸盐水泥砂浆配方中, 我们可以在 0.05 显著性水平下拒绝 $H_0: \mu_1 = \mu_2$. 结论采用这种表述方式通常是不充分的, 因为, 决策者不能从结论中知道检验统计量的计算值是恰好落在拒绝域, 还是远离拒绝域. 此外, 用这种方式表述结果是将预先指定的显著性水平强加给该信息的其他使用者. 这种方法也许并不令人满意, 因为一些决策者会对 α=0.05 所蕴含的风险感到不安.

为了避免这些困难, 实践中广泛采用**P值**方法. P 值是当零假设 H_0 为真时检验统计量至少取到像统计量观测值一样极端的概率. 因此, P 值表达了更多关于拒绝 H_0 的证据的权重信息. 所以决策者可以在任何给定显著性水平下做出结论. 形式上, 我们定义**P 值**是拒绝零假设 H_0 的最小显著性水平.

当零假设 H_0 被拒绝时, 通常称检验统计量 (和数据) 是显著的, 因而我们可以认为 P 值是数据显著时的最小 α 水平. 一旦 P 值已知, 决策者可以判定数据是如何显著的, 而不必受制于数据分析师预先选定的显著性水平.

为检验计算精确的 P 值并不总是很容易. 但是, 大多数用于统计分析的现代计算机程序都可以输出 P 值, 一些计算器可能也具有这种功能. 下面我们举例说明如何近似计算硅酸盐水泥

砂浆实验的 P 值. 因为 $|t_0| = 2.20 > t_{0.025,18} = 2.101$, 从而 P 值小于 0.05. 从附录表 II 可以看到, 对于自由度为 18、尾部概率为 0.01 时的 t 分布, $t_{0.01,18} = 2.552$, 现在 $|t_0| = 2.20 < 2.552$, 又因为备择假设是双边的, 所以知道 P 值一定在 0.05 和 $2\times(0.01) = 0.02$ 之间. 一些计算器也有计算 P 值的功能. 比如, 计算器 HP-48 就具有这种功能. 由这个计算器可得, 硅酸盐水泥砂浆实验的 P 值在 $t_0 = -2.20$ 时 $P= 0.041\ 1$. 所以, 零假设 $H_0: \mu_1 = \mu_2$ 在显著性水平 $\alpha \geqslant 0.041\ 1$ 时将被拒绝.

5. 计算机处理

许多统计软件包都具有统计检验的功能. 表 2.2 表示的就是应用在硅酸盐水泥砂浆实验的两样本 t 检验方法的 Minitab 输出. 注意输出包括了两个样本的一些描述性统计量. (缩写 "SE mean" 指均值的标准误, $s/\sqrt{n}$) 和一些关于两个均值差异的置信区间信息 (我们将在 2.4.3 节和 2.5 节中讨论). 这个程序也能检验感兴趣的假设, 通过分析师给定备择假设的种类 ("not =" 意味着 $H_1: \mu_1 \neq \mu_2$) 和给定的 α 的水平 (这里 $\alpha= 0.05$).

表 2.2 Minitab 中的两样本 t 检验

```
Two-sample T for Modified vs Unmodified
                 N        Mean        StDev        SE Mean
Modified        10      16.764        0.316           0.10
Unmodified      10      17.042        0.248          0.078

Difference=mu (Modified)-mu (Unmodified)
Estimate for difference:  -0.278000
95% CI for difference:  (-0.545073, -0.010927)
T-Test of difference=0 (vs not=):  T-value=-2.19
P-Value=0.042 DF=18
Both use Pooled StDev=0.2843
```

输出包括 t_0 的计算值、P 值 (称为显著性水平) 和根据给定的 α 值做出的决策. 注意 t 统计量的计算值与通常手算的值稍有不同, P =0.042. 许多软件包并不能输出小于一些预定值 (比如 0.000 1) 的实际 P 值, 相反它们会返回像 "< 0.001" 的 "默认值", 或者在某些情形中返回零.

6. t 检验中的假定检验

在使用 t 检验方法中, 我们假定, 所有的样本都是从正态分布的独立总体中得到的, 所有总体的标准差或方差都是相等的, 而且观测都是独立的随机变量. 独立性假设是关键的, 如果实验次序是随机的 (如果可以的话, 其他实验单元和材料的选择也是随机的), 这个条件通常是可以满足的. 利用**正态概率图**很容易检验方差齐性和正态性假设.

一般地, 概率图是基于主观可见的实验数据来判断样本数据是否服从假设分布的一种作图技巧. 方法非常简单, 大多数统计软件包都可以很快完成. **补充材料**讨论了正态概率图的人工构造.

为了构造正态概率图, 首先将样本观测值从小到大进行排序. 即样本 $y_1, y_2, \cdots, y_n$ 被排成 $y_{(1)}, y_{(2)}, \cdots, y_{(n)}$, 其中, $y_{(1)}$ 是观测值中最小的, $y_{(2)}$ 是观测值中次小的, 依此类推, $y_{(n)}$ 是最大的. 次序观测值 $y_{(j)}$ 根据它们的累积频率 $(j-0.5)/n$ 来作图, 标注好累积频率刻度, 如果假设的分布足够描述该数据, 所画的点将近似地落在一条直线上; 如果所画的点明显偏离任意一条直线, 则假设的模型是不合适的. 一般来讲, 判断数据点是否落在一条直线上是主观的.

为了说明这种方法, 假定要检验硅酸盐水泥砂浆实验的粘结抗拉强度服从正态分布这个假设. 首先只考虑未改良配方的观测值. 图 2.11 所示为计算机生成的正态概率图. 大多数正态概率图都在左面垂直刻度上列出 $100(j-0.5)/n$(偶尔在右面垂直刻度上列出 $100[1-(j-0.5)/n]$), 而变量值标在横坐标上. 一些计算机生成的正态概率图将累积频率转换为标准正态 z 值. 主观选择的直线可以根据图中的点画出来. 绘制这条直线时, 应该着重考虑图的中间点, 而不是两边的极端点. 一个好的规则就是在第 25 个百分位数的点和第 75 个百分位数的点之间画一条直线. 这就是图 2.11 中各样本直线的确定办法. 在估计直线的 "靠近" 的点时, 想象沿直线有一支粗铅笔. 如果所有的点都被这支想象的铅笔所覆盖, 则可以认为这些数据服从正态分布. 因为图 2.11 中各样本的点都通过了粗铅笔的检测, 我们可以得出结论, 即对于改良和未改良泥浆的粘结抗拉强度, 正态分布都是合适的模型.

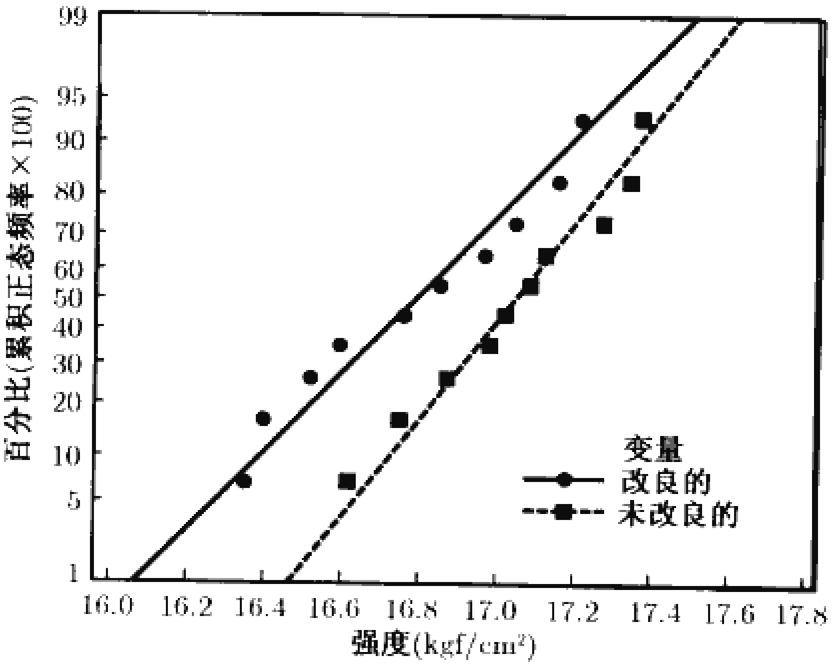


图 2.11 硅酸盐水泥砂浆实验中的粘结抗拉强度的正态概率图

可以直接通过正态概率图得到均值和标准差的估计. 均值用概率图的 50%分位点来估计, 标准差用 84%和 50%分位点的差来估计. 这意味着, 我们可以通过简单比较图 2.11 的两条直线的斜率, 从而证实硅酸盐水泥砂浆实验的总体方差相等的这一假设. 若两条直线有类似的斜率, 则方差相等的假设是合理的. 如果偏离了假设, 则可以采用 2.4.4 节中描述的 t 检验方法. 补充材料中有更多关于 t 检验的假设检验的内容.

若严重偏离假设, t 检验的效果将会受到影响. 一般地, 我们不是很介意与假设偏离不太严重的情况, 但是不应当忽视任何独立性假设的不成立和强烈的非正态信息. 检验的显著性水平和检测均值之间差异的能力反过来将会受到背离假设的影响. 变换是处理这种问题的一种方法. 第 3 章将会更详细地讨论它. 如果观测值是来自非正态总体, 可以应用非参数检验方法. 详细的内

容见 Montgomery and Runger(2003).

7. 关于 t 检验的另一个说明

我们刚才提出的两样本 t 检验法理论上依赖于下述假定, 即从其中随机选取样本的两个总体都是正态的. 虽然正态性假定对形式上推导检验程序是必需的, 但正如前面讨论的, 适度地背离正态性的假定也不会严重影响结论. 可以论证, 利用随机化设计可以检验假设而无需对分布的形式作任何假定. 简要说来, 理由如下. 如果这些处理不起作用, 则由 20 个观测值能产生的所有的 [20! /(10! 10!)]=184 756 个可能的方式都具有相同的发生可能性. 这 184 756 个可能结果中的每一个都对应着 t_0 的一个值. 如果把由这些数据而实际得出的 t_0 值与这 184 756 个可能值的集合对照之后发现它不寻常地大或不寻常地小, 则表明 $\mu_1 \neq \mu_2$.

这种过程称为**随机化检验**. 可以证明, t 检验法是随机化检验的一个良好近似. 于是, 我们使用 t 检验法 (以及其他可以作为随机化检验的近似方法) 时, 无需顾及关于正态性的假定. 这也正是像正态概率图这样的简单方法就足以检验正态假定的原因.

2.4.2 样本量的选取

选取恰当的样本量是任何实验设计问题最重要的方面之一. 样本量的选择与第 II 类错误概率 β 密切相关. 假如要检验假设

$$H_0: \mu_1 = \mu_2, \qquad H_1: \mu_1 \neq \mu_2$$

且均值不相等, 令 $\delta = \mu_1 - \mu_2$. 因为 $H_0: \mu_1 = \mu_2$ 不真, 我们关心未能拒绝 H_0 的错误. 第 II 类错误的概率依赖于均值的实际差别 δ. 对于一个具体的样本量, β 与 δ 的图像叫做检验法的**抽检特性曲线**(operating characteristic curve) 或**OC曲线**. β 还是样本量的函数. 一般说来, 给定一个 δ 值, β 随着样本量的增大而减小. 也就是说, 大样本量比小样本量更容易检测出一个给定的均值差.

对于两个总体的方差 σ_1^2 与 σ_2^2 未知但相等 $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$, 显著性水平为 $\alpha = 0.05$ 的情形, 假设

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

的一族抽检特性曲线如图 2.12 所示. 还假定这些曲线来自两个样本量相等的总体, 亦即 $n_1 = n_2 = n$. 图 2.12 中水平轴的参数是

$$d = \frac{|\mu_1 - \mu_2|}{2\sigma} = \frac{|\delta|}{2\sigma}$$

用 2σ 去除 $|\delta|$ 便于实验者利用同一族曲线而无需顾及方差的值 (均值差表示为标准差的单位数). 而且, 用于构造这些曲线的样本量实际上就是 $n^* = 2n - 1$.

从这些曲线中, 我们注意到以下几点.

(1) 已知样本量和 α 时, 均值差 $\mu_1 - \mu_2$ 愈大, 第 II 类错误的概率愈小. 也就是说, 给定样本量和 α 时, 这一检验法对大的均值差比小的均值差更容易检测出来.

(2) 已知均值差和 α 时, 样本量愈大, 第 II 类错误概率愈小. 即, 要检测一个给定的均值差 δ, 增大样本量将使检验方法更为有效.

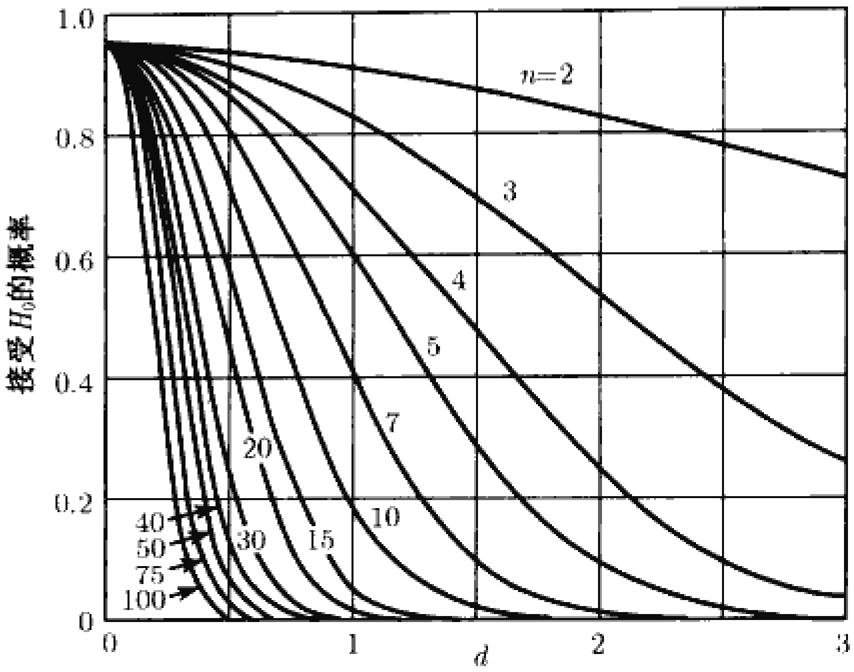


图 2.12 $\alpha = 0.05$ 时双边 t 检验的抽检特性曲线

(承蒙惠允, 本图复制自 Operating Characteristics for the Common Statistical Tests of Significance, C. L. Ferris, F. E. Grubbs, and G. L. Weaver, Annals of Mathematical Statistics, June 1946)

抽检特性曲线常有助于选取实验中的样本量. 例如, 考虑前面讨论过的硅酸盐水泥砂浆问题. 设想两种配方的平均强度相差为 0.5 kgf/cm^2, 该值通常会影响砂浆使用. 如果均值差异值至少这么大, 我们将以一个较大的概率检测它. 由此, 因为 $\mu_1 - \mu_2 = 0.5$ kgf/cm^2 是我们希望检测的均值差的 "临界" 值, 所以得到图 2.12 中抽检特性曲线水平轴的参数 d 是

$$d = \frac{|\mu_1 - \mu_2|}{2\sigma} = \frac{0.5}{2\sigma} = \frac{0.25}{\sigma}$$

不幸的是, d 含有未知数 σ. 但是, 假设我们根据以前的经验, 认为强度的观测值的标准差超过 0.25 kgf/cm^2 是不大可能发生的. 那么, 在上式中对 d 用 $\sigma = 0.25$ 就得出 d=1. 如果当 $\mu_1 - \mu_2 = 0.5$ 时我们希望以 95%的可能性拒绝零假设的话, 则 $\beta = 0.05$, 由图 2.12 知, 当 $\beta = 0.05$ 且 d=1 时, 近似地得到 $n^* = 16$. 因为 $n^* = 2n - 1$, 所以需要的样本量是

$$n = \frac{n^* + 1}{2} = \frac{16 + 1}{2} = 8.5 \approx 9$$

从而我们采用的样本量为 $n_1 = n_2 = n = 9$.

在上例中, 实验者实际上用的样本量为 10. 实验者选择稍大点的样本量也许是针对一种可能性, 即共同标准差 σ 的先验估计太保守了, 它有时可能会大于 0.25.

抽检特性曲线在实验设计问题中对选取样本量常常起着重要的作用. 这方面的内容在以后几章中都会讨论到. 在其他简单比较实验中使用抽检特性曲线的讨论, 类似于两样本 t 检验法, 请参阅 Montgomery and Runger(2003).

许多统计软件包可以辅助实验者进行势和样本量的计算. 下面的方框演示了 Minitab 中的关于硅酸盐水泥砂浆问题的两样本 t 检验的势和样本量的几个计算. 输出的第一部分再次给出了 OC 曲线的分析, 假定强度的标准差为 0.25 kgf/cm^2, 计算临界均值差为 0.5 kgf/cm^2 所必需的样本量. Minitab 中的答案为 $n_1 = n_2$ =8, 相当接近 OC 曲线分析的结果. 输出的第二部分计算了本例的势, 这里临界均值差非常小, 只有 0.25 kgf/cm^2. 势下降得很快, 从 0.95 到 0.562.

最后一部分确定了实际均值差为 0.25 kgf/cm^2, 势至少为 0.9 所必需的样本量. 算出的样本量相对较大, $n_1 = n_2 = 23$.

```
Power and sample size

2-Sample t Test
Testing mean 1=mean 2(versus not=)
Calculating power for mean 1=mean 2+difference
Alpha=0.05 Sigma=0.25

                      Sample            Target            Actual
Difference              Size             Power             Power
0.5                        8            0.9500            0.9602

Power and Sample Size
2-Sample t Test

Testing mean 1=mean 2(versus not=)
Calculating power for mean 1=mean 2+difference
Alpha=0.05     Sigma=0.25

                      Sample
Difference              Size             Power
0.25                      10            0.5620

Power and Sample Size
2-Sample t Test

Testing mean 1=mean 2(versus not =)
Calculating power for mean 1=mean 2+difference
Alpha=0.05  Sigma =0.25

                      Sample            Target            Actual
Difference              Size             Power             Power
0.25                      23            0.9000            0.9125
```

2.4.3 置信区间

尽管假设检验是一种有用的方法, 但它有时也不能说出全部细节. 人们喜欢提供一个区间, 期待问题中的参数或参数值落在这一区间之内. 这种区间称为**置信区间**(confidence interval). 在很多工程和工业实验中, 实验者已经知道均值 μ_1 与 μ_2 有差别, 因此对于 $\mu_1 = \mu_2$ 的假设检验就不大有兴趣. 实验者通常更感兴趣的是均值差 $\mu_1 - \mu_2$ 的置信区间.

设 θ 是一未知参数, 我们来确定一个置信区间. 为了得出 θ 的一个区间估计, 需要找出两个

统计量 L 和 U, 使得概率命题

$$\mathrm{P}(L \leqslant \theta \leqslant U) = 1 - \alpha \tag{2.27}$$

为真. 区间

$$L \leqslant \theta \leqslant U \tag{2.28}$$

叫做参数 θ 的**100(1 − α)%置信区间**. 这个区间的意思是, 在重复的随机抽样中, 如果得出很多这样的区间, 则其中的 $100(1-\alpha)\%$会含有 θ 的真值. 统计量 L 与 U 分别叫做**下置信限**(lower confidence limit) 和**上置信限**(upper confidence limit), $1-\alpha$ 叫做**置信系数**(confidence coefficient). 如果 $\alpha = 0.05$, 则 (2.28) 式叫做 θ 的 95%置信区间. 置信区间有一个频率性的解释, 即, 我们不知道对于这个特定样本说来命题是否为真, 但我们知道, 被用来产生置信区间的方法有 $100(1-\alpha)\%$次得出正确的命题.

对于硅酸盐水泥问题, 设我们要对真实的均值差 $\mu_1 - \mu_2$ 找一个 $100(1-\alpha)\%$的置信区间. 此区间可用下述方法导出. 统计量

$$\frac{\overline{y}_1 - \overline{y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

服从 $t_{n_1+n_2-2}$ 分布. 于是

$$\mathrm{P}\left[-t_{\alpha/2,n_1+n_2-2} \leqslant \frac{\overline{y}_1 - \overline{y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leqslant t_{\alpha/2,n_1+n_2-2}\right] = 1 - \alpha$$

$$\mathrm{P}\Big(\overline{y}_1 - \overline{y}_2 - t_{\alpha/2,n_1+n_2-2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leqslant \mu_1 - \mu_2$$

$$\leqslant \overline{y}_1 - \overline{y}_2 + t_{\alpha/2,n_1+n_2-2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\Big) = 1 - \alpha \tag{2.29}$$

比较 (2.29) 式与 (2.27) 式, 可以看出

$$\overline{y}_1 - \overline{y}_2 - t_{\alpha/2,n_1+n_2-2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leqslant \mu_1 - \mu_2 \leqslant \overline{y}_1 - \overline{y}_2 + t_{\alpha/2,n_1+n_2-2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{2.30}$$

是 $\mu_1 - \mu_2$ 的一个 $100(1-\alpha)\%$置信区间.

代入 (2.30) 式就得到 95%置信区间估计如下:

$$16.76 - 17.04 - (2.101)0.284\sqrt{\frac{1}{10} + \frac{1}{10}} \leqslant \mu_1 - \mu_2 \leqslant 16.76 - 17.04 + (2.101)0.284\sqrt{\frac{1}{10} + \frac{1}{10}}$$

$$-0.28 - 0.27 \leqslant \mu_1 - \mu_2 \leqslant -0.28 + 0.27$$

$$-0.55 \leqslant \mu_1 - \mu_2 \leqslant -0.01$$

于是, 均值差的 95%置信区间是从 -0.55 kgf/cm^2 到 -0.01 kgf/cm^2. 换一种说法, 置信区间是 $\mu_1 - \mu_2 = -0.28$ kgf/cm$^2 \pm 0.27$ kgf/cm^2, 或强度均值的差异为 -0.28 kgf/cm^2. 注意, 因为 $\mu_1 - \mu_2 = 0$不包含在此区间中, 所以, 所得数据在 5%的显著性水平上不支持 $\mu_1 = \mu_2$ 的假设.

(两样本 t 检验的 P 值为 0.042, 稍小于 0.05). 未改良配方的平均强度超过改良配方的平均强度是很可能的. 注意表 2.2, Minitab 在假设检验方法操作中也得到了这一置信区间.

2.4.4 $\sigma_1^2 \neq \sigma_2^2$ 的情形

如果检验

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

而不能合理地假定方差 σ_1^2 与 σ_2^2 相等, 则两样本 t 检验法须要稍加修正. 这时检验统计量为

$$t_0 = \frac{\overline{y}_1 - \overline{y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{2.31}$$

这个统计量不再精确地服从 t 分布. 然而, 如果用

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}} \tag{2.32}$$

作为自由度, 则 t_0 的分布相当接近于 t 分布. 这种 t 检验的方法适用于正态概率图表明的方差不等的情形. 可以容易地推出一个公式来找出不等方差情形的均值差异的置信区间.

2.4.5 σ_1^2 与 σ_2^2 为已知的情形

如果两个总体的方差都是**已知**的, 则检验假设

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

可以用统计量

$$z_0 = \frac{\overline{y}_1 - \overline{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{2.33}$$

如果两个总体都是正态的, 或者样本量足够大, 以至于可以用中心极限定理, 则当零假设为真时, z_0 的分布是 $N(0,\ 1)$. 这样, 我们用正态分布而不用 t 分布来求拒绝域. 特别地, 当 $|z_0| > z_{\alpha/2}$ 时我们就拒绝 H_0, 其中 $z_{\alpha/2}$ 是标准正态分布的上 $\alpha/2$ 百分位数.

与前面几节的 t 检验法不同, 已知方差时检验均值不需要假定从正态总体中抽样. 可以用中心极限定理来证明样本均值差 $\overline{y}_1 - \overline{y}_2$ 是近似正态分布的.

方差已知时关于 $\mu_1 - \mu_2$ 的 $100(1-\alpha)\%$的置信区间是

$$\overline{y}_1 - \overline{y}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leqslant \mu_1 - \mu_2 \leqslant \overline{y}_1 - \overline{y}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{2.34}$$

正如前面所注明的那样, 置信区间常常是假设检验方法的一个有用的补充.

2.4.6 均值与已知值的比较

有些实验仅比较一个总体的均值 μ 和一个已知值 μ_0. 这一假设是

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

如果总体是正态的, 其方差已知, 或者总体是非正态的, 但样本量足够大以至于可以用中心极限定理, 则这一假设可以直接用正态分布来检验. 检验统计量是

$$z_0 = \frac{\overline{y} - \mu_0}{\sigma/\sqrt{n}} \tag{2.35}$$

如果 $H_0: \mu = \mu_0$ 为真, 则 z_0 的分布是 $N(0, 1)$. 因此, 对 $H_0: \mu = \mu_0$ 的判别法则是, 如果 $|z_0| > z_{\alpha/2}$, 则拒绝零假设. 在零假设中均值的已知值 μ_0 通常可由三种方法决定. 可以根据过去的论证、知识或经验而得到; 可以是描述所研究情况的某种理论或模型的结果; 也可以是合同规定的结果.

真实总体均值的 $100(1-\alpha)\%$置信区间是

$$\overline{y} - z_{\alpha/2}\sigma/\sqrt{n} \leqslant \mu \leqslant \overline{y} + z_{\alpha/2}\sigma/\sqrt{n} \tag{2.36}$$

例 2.1 一供应商向一纺织厂提供一批纤维. 工厂需要了解这批纤维的平均抗断强度是否超过 200 psi. 如果超过 200 psi, 工厂就接收这些纤维. 过去的经验表明, 抗断强度方差的一个合理值是 100 (psi)^2. 需要检验的假设是

$$H_0: \mu = 200, \qquad H_1: \mu > 200$$

注意这是单边备择假设. 于是, 仅当零假设 $H_0: \mu = 200$ 被拒绝时 (即, 若 $z_0 > z_\alpha$), 才接收这批纤维.

随机选取 4 个试件, 观测到的平均抗断强度为 $\overline{y} = 214$ psi. 检验统计量的值是

$$z_0 = \frac{\overline{y} - \mu_0}{\sigma/\sqrt{n}} = \frac{214 - 200}{10/\sqrt{4}} = 2.80$$

如果指定第 I 类错误为 $\alpha = 0.05$, 从附录的表 1 中查得 $z_\alpha = z_{0.05} = 1.645$. 这样, H_0 被拒绝, 我们得出的结论是, 这批纤维的平均抗断强度超过 200 psi.

如果总体的方差未知, 就必须加上总体服从正态分布的这条假定, 尽管适度的偏离正态性并不会严重影响所得的结论.

在方差未知的情况下, 要检验 $H_0: \mu = \mu_0$, 则要用样本方差 S^2 来估计 σ^2. 在 (2.35) 式中以 S 代替 σ, 得检验统计量

$$t_0 = \frac{\overline{y} - \mu_0}{S/\sqrt{n}} \tag{2.37}$$

当 $|t_0| > t_{\alpha/2,n-1}$ 时, 拒绝零假设 $H_0: \mu = \mu_0$, 其中 $t_{\alpha/2,n-1}$ 是自由度为 $n-1$ 的 t 分布的上 $\alpha/2$ 百分位数. 此时的 $100(1-\alpha)\%$置信区间为

$$\overline{y} - t_{\alpha/2,n-1}S/\sqrt{n} \leqslant \mu \leqslant \overline{y} + t_{\alpha/2,n-1}S/\sqrt{n} \tag{2.38}$$

2.4.7 小结

表 2.3 和表 2.4 简要地列出了上面关于样本均值所讨论的检验方法. 对双边和单边备择假设都列出了拒绝域.

表 2.3 方差已知时关于均值的检验

假设	检验统计量	拒绝域	假设	检验统计量	拒绝域
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$\|z_0\| > z_{\alpha/2}$	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$		$\|z_0\| > z_{\alpha/2}$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$z_0 = \frac{\overline{y}-\mu_0}{\sigma/\sqrt{n}}$	$z_0 < -z_\alpha$	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$z_0 = \frac{\overline{y}_1-\overline{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$z_0 < -z_\alpha$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$z_0 > z_\alpha$	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$		$z_0 > z_\alpha$

表 2.4 方差未知时关于正态分布均值的检验

假设	检验统计量	拒绝域
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$\|t_0\| > t_{\alpha/2,n-1}$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$t_0 = \frac{\overline{y}-\mu_0}{S/\sqrt{n}}$	$t_0 < -t_{\alpha,n-1}$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$t_0 > t_{\alpha,n-1}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	若 $\sigma_1^2 = \sigma_2^2$ $t_0 = \frac{\overline{y}_1-\overline{y}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $v = n_1 + n_2 - 2$	$\|t_0\| > t_{\alpha/2,v}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	若 $\sigma_1^2 \neq \sigma_2^2$ $t_0 = \frac{\overline{y}_1-\overline{y}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$	$t_0 < -t_{\alpha,v}$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$v = \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1}+\frac{(S_2^2/n_2)^2}{n_2-1}}$	$t_0 > t_{\alpha,v}$

2.5 关于均值差的推断, 配对比较设计

2.5.1 配对比较问题

在有些简单比较实验中, 将实验材料进行配对比较能极大地提高估计的精确度. 例如, 考虑一台硬度试验机, 它以恒力把杆的尖端压头压入一块金属试件, 然后通过测量压头压入的深度, 即可得知试件的硬度. 这台机器有两根不同的压头, 虽然两根压头产生的测量精度 (变异性) 似乎相同, 但仍怀疑一根压头得到的硬度读数与另一根压头所得到的硬度读数是不同的.

实验方法如下. 随机选取一批 (例如, 20 块) 金属试件. 这些试件的一半用压头 1 检测, 另一半用压头 2 检测. 试件对压头的具体分配是随机的. 因为这是一种完全随机化设计, 所以两个样本的平均硬度可以用 2.4 节所述的 t 检验法来比较.

这个问题多少可以反映出完全随机化设计的一个严重缺陷. 设金属试件是由不同的条材切割而来, 而这些条材是在不同炉次或可能对硬度有影响的不尽相同的方式下生产的. 由于试件之间缺乏均匀性, 这就会产生硬度数据的变异性, 并会增大实验的误差, 从而使得压头之间的真实差别难以检测出来.

针对这种可能性, 我们考虑另一种实验设计. 假定每一块试件足够大, 使得在同一块试件可以进行两次硬度测定. 这种设计是把每一块试件划分为两部分, 然后对每块试件的一半随机地指定一根压头, 对余下的一半用另一根压头. 对每一块指定的试件, 压头实验的顺序也是随机选择的. 按照这种设计对 10 块试件进行实验, 测得的 (规范化) 数据如表 2.5 所示.

表 2.5 硬度检验实验的数据

试 件	压头 1	压头 2	试 件	压头 1	压头 2
1	7	6	6	3	2
2	3	3	7	2	4
3	3	5	8	9	9
4	4	3	9	5	4
5	8	8	10	4	5

描述这种实验数据的**统计模型**(statistical model) 为

$$y_{ij} = \mu_i + \beta_j + \varepsilon_{ij} \begin{cases} i = 1, 2 \\ j = 1, 2, \cdots, 10 \end{cases} \tag{2.39}$$

其中 y_{ij} 是压头 i 在试件 j 上的硬度观测值, μ_i 是第 i 根压头的平均硬度读数的真值, β_j 是第 j 块试件的硬度效应, ε_{ij} 是随机实验误差, 其均值为零、方差为 σ_i^2. 即, σ_1^2 是用压头 1 测硬度时测量值的方差, σ_2^2 是用压头 2 测硬度时测量值的方差.

如果计算第 j 个配对差

$$d_j = y_{1j} - y_{2j}, \quad j = 1, 2, \cdots, 10 \tag{2.40}$$

则此差的期望值是

$$\mu_d = \mathrm{E}(d_j) = \mathrm{E}(y_{1j} - y_{2j}) = \mathrm{E}(y_{1j}) - \mathrm{E}(y_{2j}) = \mu_1 + \beta_j - (\mu_2 + \beta_j) = \mu_1 - \mu_2$$

即, 我们可以利用关于硬度读数差的均值 μ_d 的推断来作出关于两根压头的硬度读数均值差 $\mu_1 - \mu_2$ 的推断. 当观测值用这种方式配对时, 试件的附加效应 β_j 会消失.

检验 $H_0: \mu_1 = \mu_2$ 等价于检验

$$H_0: \mu_d = 0, \qquad H_1: \mu_d \neq 0$$

这一假设的检验统计量是

$$t_0 = \frac{\overline{d}}{S_d/\sqrt{n}} \tag{2.41}$$

其中

$$\overline{d} = \frac{1}{n}\sum_{j=1}^{n} d_j \tag{2.42}$$

是配对差的样本均值, 且

$$S_d = \left[\frac{\sum_{j=1}^{n}(d_j - \overline{d})^2}{n-1}\right]^{1/2} = \left[\frac{\sum_{j=1}^{n} d_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} d_j\right)^2}{n-1}\right]^{1/2} \tag{2.43}$$

是配对差的样本标准差. 当 $|t_0| > t_{\alpha/2,n-1}$ 时, 则拒绝 $H_0: \mu_d = 0$. 因为每个实验单元中因子水平的观测是成对出现的, 所以这种方法通常叫做 "**配对 t 检验**".

由表 2.5 的数据得

$$d_1 = 7-6=1,\ d_2 = 3-3=0,\ d_3 = 3-5=-2,\ d_4 = 4-3=1,\ d_5 = 8-8=0,$$
$$d_6 = 3-2=1,\ d_7 = 2-4=-2,\ d_8 = 9-9=0,\ d_9 = 5-4=1,\ d_{10} = 4-5=-1$$

于是

$$\overline{d} = \frac{1}{n}\sum_{j=1}^{n} d_j = \frac{1}{10}(-1) = -0.10$$

$$S_d = \left[\frac{\sum_{j=1}^{n} d_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} d_j\right)^2}{n-1}\right]^{1/2} = \left[\frac{13 - \frac{1}{10}(-1)^2}{10-1}\right]^{1/2} = 1.20$$

假定选择 $\alpha = 0.05$. 为了作决策, 我们将计算 t_0, 如果 $|t_0| > t_{0.025,9} = 2.262$, 则拒绝 H_0.

配对 t 检验统计量的计算值是

$$t_0 = \frac{\overline{d}}{S_d/\sqrt{n}} = \frac{-0.10}{1.20/\sqrt{10}} = -0.26$$

因为 $|t_0| = 0.26 < t_{0.025,9} = 2.262$, 所以我们不能拒绝假设 $H_0: \mu_d = 0$. 也就是说, 没有证据来证明两根压头会得出不同的硬度读数. 图 2.13 所示是自由度为 9 的 t_0 分布, 它是本次 t 检验的参考分布, 并给出了相对于拒绝域的 t_0 的值.

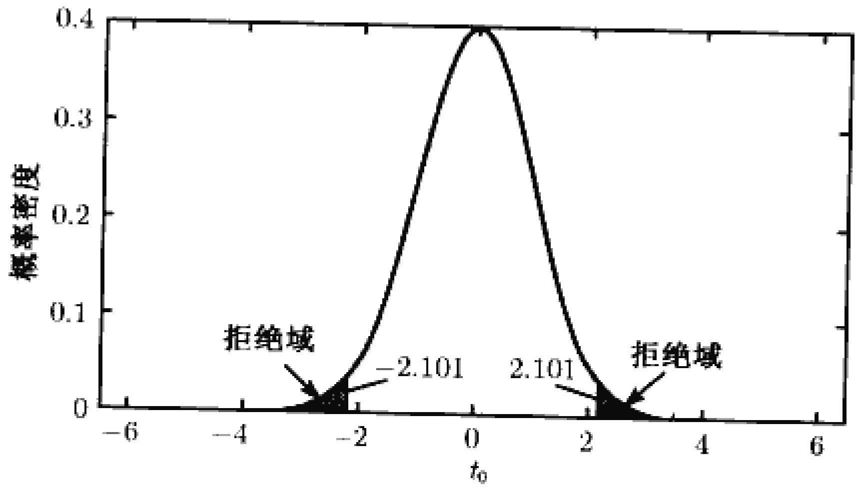


图 2.13 硬度检验问题的参考分布 (自由度为 9 的 t 分布)

表 2.6 显示了该问题的 Minitab 配对 t 检验程序的计算机输出. 注意到检验的 P 值是 $P \approx 0.80$, 这意味着在任何合理的显著性水平下, 我们都不能拒绝零假设.

2.5.2 配对比较设计的优点

实际用于这一实验的设计叫做**配对比较设计**(paired comparison design), 这也解释了在 1.3 节中讨论的区组化原则. 实际上, 它是更一般的**随机化区组设计**的设计类型的特殊情况. 区组一

表 2.6 硬度实验问题的配对 t 检验 Minitab 输出

```
Paired T for Tip 1-Tip 2

                    N          Mean         StDev        SE  Mean
Tip 1              10         4.800         2.394           0.757
Tip 2              10         4.900         2.234           0.706
Difference         10        -0.100         1.197           0.379

95% CI for mean difference:  (-0.956, 0.756)
t-Test of mean difference=0(vs not=0):
T-Value=-0.26     P-VaLue=0.798
```

词理解为一个相对均匀的实验单元 (在我们的例子中, 金属试件就是区组), 并且区组表示对完全随机化的一个约束, 因为处理组合仅仅在区组内部是随机化的. 我们可查阅第 4 章中所叙述的这类设计, 在那一章中, 设计的数学模型 [即 (2.39) 式] 在写法上会稍有不同.

在结束这个实验之前, 有几点需要说明. 注意, 虽然我们取了 $2n$=2(10)=20 个观测值, 但对 t 统计量只用了 $n-1$=9 个自由度. (我们知道, t 的自由度增加时, 检验法会更加灵敏.) 区组化或配对使我们 "失去" 了 $n-1$ 个自由度, 但我们希望, 通过消除额外的变异性来源 (试件之间的差别), 我们能得到更好的信息.

通过比较配对差的样本标准差 S_d 与合并的样本标准差 S_p, 我们可以从配对设计中得到一个信息质量的指标. 这里的合并标准差 S_p 是我们设想表 2.5 的数据是按完全随机化方式进行的实验所得到的. 把表 2.5 的数据作为两个独立的样本, 由 (2.25) 式算得的合并标准差是 $S_p=2.32$. 将这一数值与 $S_d=1.20$ 比较, 即可看出区组化或配对将变异性的估计几乎减少了 50%.

当我们真正需要而又没有区组化 (或将观测值配对) 时, S_p 通常比 S_d 大得多. 很容易说明这一点. 如果将观测值组配对, 容易说明 S_d^2 在 (2.39) 式给出的模型下是差值 d_j 的方差的无偏估计, 因为计算差值时区组因子 (β_j) 被抵消了. 但是, 如果我们没有区组 (或配对), 则将观测值视为两个独立样本, 此时在 (2.39) 式给出的模型下, S_d^2 不能视为 σ^2 的无偏估计. 事实上, 假定所有总体方差相等,

$$E(S_p^2)=\sigma^2+\sum_{j=1}^{n}\beta_j^2$$

也就是, 区组效应 β_j 使方差估计偏高了. 这就是为什么区组适合作为减少噪声设计方法的原因.

我们也可以把这一信息用关于 $\mu_1-\mu_2$ 的置信区间来表示. 利用配对数据, 关于 $\mu_1-\mu_2$ 的一个 95%置信区间是

$$\overline{d}\pm t_{0.025,9}S_d/\sqrt{n}=-0.10\pm(2.262)(1.20)/\sqrt{10}=-0.10\pm0.86$$

反之, 用合并的或独立的分析法, 关于 $\mu_1-\mu_2$ 的一个 95%置信区间是

$$\overline{y}_1-\overline{y}_2\pm t_{0.025,18}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}=4.80-4.90\pm(2.101)(2.32)\sqrt{\frac{1}{10}+\frac{1}{10}}=-1.10\pm2.18$$

显然, 由配对分析所得的置信区间小于由独立分析所得的置信区间. 这再一次说明了区组化的噪音减小性质.

区组化并不总是最好的设计策略. 如果区组内部的变异性和区组之间的变异性相同, 则 $\bar{y}_1 - \bar{y}_2$ 的方差不管用什么样的设计都是相同的. 实际上, 在这种情况下, 区组化将是一个较差的设计选择, 因为区组化损失了 $n-1$ 个自由度, 并且实际上导致了一个关于 $\mu_1 - \mu_2$ 的较宽的置信区间. 第 4 章将对区组化作进一步的讨论.

2.6 正态分布方差的推断

在很多实验中, 我们想要知道两种处理的平均响应的可能差别. 不过, 在某些实验中, 比较数据中的变异性更重要. 例如, 在食品和饮料工业中, 重要的是装料设备的变异性要很小, 以便所有的包装都接近于额定的净重或容量. 在化学实验室中, 需要比较两种分析方法的变异性. 现在, 我们简要地讨论一下正态分布方差的假设检验和置信区间. 与检验均值不同, 检验方差的方法对正态性假定是很敏感的. 一个很好的关于正态性假定的讨论见 Davies(1956) 的附录 2A.

假如要检验假设: 一个正态总体的方差等于常数, 例如, 等于 σ_0^2. 规范地说, 就是要检验

$$H_0: \sigma^2 = \sigma_0^2, \qquad H_1: \sigma^2 \neq \sigma_0^2 \tag{2.44}$$

(2.44) 式的检验统计量是

$$\chi_0^2 = \frac{SS}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2} \tag{2.45}$$

其中 $SS = \sum_{i=1}^{n}(y_i - \bar{y})^2$是样本观测值的校正平方和. χ_0^2 的合理推断分布是自由度为 $n-1$ 的卡方分布. 当 $\chi_0^2 > \chi_{\alpha/2,n-1}^2$ 或当 $\chi_0^2 < \chi_{1-(\alpha/2),n-1}^2$ 时, 则拒绝零假设, 其中 $\chi_{\alpha/2,n-1}^2$ 与 $\chi_{1-(\alpha/2),n-1}^2$ 分别是自由度为 $n-1$ 的卡方分布的上 $\alpha/2$ 和下 $1-(\alpha/2)$ 百分位数. 表 2.7 给出单边备择假设的拒绝域. σ^2 的 $100(1-\alpha)\%$置信区间是

$$\frac{(n-1)S^2}{\chi_{\alpha/2,n-1}^2} \leqslant \sigma^2 \leqslant \frac{(n-1)S^2}{\chi_{1-(\alpha/2),n-1}^2} \tag{2.46}$$

表 2.7 关于正态分布方差的检验

假 设	检验统计量	拒绝域	假 设	检验统计量	拒绝域
$H_0:\sigma^2=\sigma_0^2$ $H_1:\sigma^2\neq\sigma_0^2$		$\chi_0^2>\chi_{\alpha/2,n-1}^2$ 或 $\chi_0^2<\chi_{1-\alpha/2,n-1}^2$	$H_0:\sigma_1^2=\sigma_2^2$ $H_1:\sigma_1^2\neq\sigma_2^2$	$F_0=\frac{S_1^2}{S_2^2}$	$F_0>F_{\alpha/2,n_1-1,n_2-1}$ 或 $F_0<F_{1-\alpha/2,n_1-1,n_2-1}$
$H_0:\sigma^2=\sigma_0^2$ $H_1:\sigma^2<\sigma_0^2$	$\chi_0^2=\frac{(n-1)S^2}{\sigma_0^2}$	$\chi_0^2<\chi_{1-\alpha,n-1}^2$	$H_0:\sigma_1^2=\sigma_2^2$ $H_1:\sigma_1^2<\sigma_2^2$	$F_0=\frac{S_2^2}{S_1^2}$	$F_0>F_{\alpha,n_2-1,n_1-1}$
$H_0:\sigma^2=\sigma_0^2$ $H_1:\sigma^2>\sigma_0^2$		$\chi_0^2>\chi_{\alpha,n-1}^2$	$H_0:\sigma_1^2=\sigma_2^2$ $H_1:\sigma_1^2>\sigma_2^2$	$F_0=\frac{S_1^2}{S_2^2}$	$F_0>F_{\alpha,n_1-1,n_2-1}$

现在考虑检验两个正态总体的方差是否相等. 如果分别从总体 1 与总体 2 中取量为 n_1 与 n_2 的独立随机样本, 则针对

$$H_0: \sigma_1^2 = \sigma_2^2, \qquad H_1: \sigma_1^2 \neq \sigma_2^2 \tag{2.47}$$

的检验统计量就是样本方差的比值

$$F_0 = \frac{S_1^2}{S_2^2} \tag{2.48}$$

F_0 的合理推断分布是分子、分母自由度分别为 n_1-1 与 n_2-1 的 F 分布. 当 $F_0 > F_{\alpha/2,n_1-1,n_2-1}$ 或当 $F_0 < F_{1-(\alpha/2),n_1-1,n_2-1}$ 时, 则拒绝零假设, 其中 $F_{\alpha/2,n_1-1,n_2-1}$ 与 $F_{1-(\alpha/2),n_1-1,n_2-1}$ 表示自由度为 n_1-1 与 n_2-1 的 F 分布的上 $\alpha/2$ 与下 $1-(\alpha/2)$ 百分位数. 附录中表 IV 只给出 F 的上 α 百分位数部分, 但上、下尾点有关系式

$$F_{1-\alpha,v_1,v_2} = \frac{1}{F_{\alpha,v_2,v_1}} \tag{2.49}$$

关于多于两个方差的检验方法将在 3.4.3 节中讨论. 我们也将讨论把方差或标准差用作更一般的实验设计中的响应变量.

例 2.2 一位化学工程师研究两类测试设备的固有变异性, 这种设备是用来监视生产过程产出的. 他推测旧设备 (即类型 1) 要比新设备有较大的方差. 因此, 他想检验假设

$$H_0: \sigma_1^2 = \sigma_2^2, \qquad H_1: \sigma_1^2 > \sigma_2^2$$

于是, 他取了 $n_1 = 12$ 和 $n_2 = 10$ 的两个随机样本的观测值, 其样本方差分别是 $S_1^2 = 14.5$ 和 $S_2^2 = 10.8$, 检验统计量是

$$F_0 = \frac{S_1^2}{S_2^2} = \frac{14.5}{10.8} = 1.34$$

从附录的表 IV 中查得 $F_{0.05,11,9} = 3.10$, 所以, 不能拒绝零假设. 也就是说, 断言旧设备的方差大于新设备的方差没有充分的统计依据.

总体方差的比值 σ_1^2/σ_2^2 的 $100(1-\alpha)\%$的置信区间是

$$\frac{S_1^2}{S_2^2}F_{1-\alpha/2,n_2-1,n_1-1} \leqslant \frac{\sigma_1^2}{\sigma_2^2} \leqslant \frac{S_1^2}{S_2^2}F_{\alpha/2,n_2-1,n_1-1} \tag{2.50}$$

为说明 (2.50) 式的用法, 例 2.2 中的方差比值 σ_1^2/σ_2^2 的 95%置信区间是

$$\frac{14.5}{10.8}(0.255) \leqslant \frac{\sigma_1^2}{\sigma_2^2} \leqslant \frac{14.5}{10.8}(3.59)$$

$$0.34 \leqslant \frac{\sigma_1^2}{\sigma_2^2} \leqslant 4.82$$

此处用到了 $F_{0.025,9,11} = 3.59$ 以及 $F_{0.975,9,11} = 1/F_{0.025,11,9} = 1/3.92 = 0.255$.

2.7 思 考 题

2.1 纤维的抗断强度要求至少为 150 psi. 过去的经验表明, 抗断强度的标准差是 σ=3 psi, 检验具有 4 个试件的一个随机样本. 结果是 $y_1 = 145$, $y_2 = 153$, $y_3 = 150$, $y_4 = 147$.

(a) 叙述你认为在该实验中应检验的假设.

(b) 用 α=0.05 检验这些假设, 你的结论是什么?

(c) 确定 (b) 中检验的 P 值.

(d) 构造平均抗断强度的一个 95%置信区间.

2.2 假定在 25 ℃时液体洗涤剂的黏度平均为 800 厘泡, 随机收集 16 个批次的洗涤剂样本, 黏度均值为 812, 假定我们已知黏度的标准差是 σ=25 厘泡.

(a) 叙述应被检验的假设.

(b) 用 α=0.05 检验这些假设, 你的结论是什么?

(c) 确定本次检验的 P 值.

(d) 确定均值的一个 95%置信区间.

2.3 某工厂生产线生产的钢轴的平均直径为 0.255 英寸. 已知直径的标准差为 $\sigma = 0.000\ 1$ 英寸. 一个 10 根钢轴的随机样本, 其平均直径为 0.254 5 英寸.

(a) 建立关于均值 μ 的合理假设.

(b) 用 α=0.05 检验这些假设, 你的结论是什么?

(c) 确定本次检验的 P 值?

(d) 构造钢轴平均直径的一个 95%置信区间.

2.4 一个正态分布随机变量, 均值 μ 未知, 方差 $\sigma^2 = 9$ 已知. 确定样本量, 使之能构造一个均值的总长度为 1.0 的 95%置信区间.

*2.5 人们关心碳酸饮料的储存期限. 随机抽取 10 瓶加以检验, 得出结果如下:

天数									
108	124	124	106	115	138	163	159	134	139

(a) 我们试图证实平均储存期限大于 120 天. 建立研究这个论断的合理假设.

(b) 用 α=0.01 检验这些假设, 你的结论是什么?

(c) 确定 (b) 中检验的 P 值.

(d) 构造平均储存期限的一个 99%置信区间.

2.6 考虑思考题 2.5 的储存期限数据. 能用一个正态分布来充分描述或拟合储存期限吗? 偏离这个假定对解决思考题 2.5 的检验方法有什么影响?

2.7 修理一台电子仪器所需的时间是一个正态分布随机变量, 以小时为测量单位. 随机选取 16 台这种仪器, 其修理时间如下:

小时							
159	280	101	212	224	379	179	264
222	362	168	250	149	260	485	170

(a) 希望了解平均修理时间是否大于 225 小时. 建立研究这个论点的合理假设.

(b) 用 α=0.05 检验 (a) 中提出的假设, 你的结论是什么?

(c) 确定本次检验的 P 值.

(d) 构造平均修理时间的一个 95%置信区间.

2.8 再考虑思考题 2.7 的修理时间问题. 依据你的意见, 正态分布能否充分拟合修理时间?

2.9 有两台机器用来充装净容量为 16.0 盎司的塑料瓶. 假定充装容量是正态的, 其标准差为 $\sigma_1 = 0.015$ 和 $\sigma_2 = 0.018$. 质量管理部门怀疑那两台机器是否充装同样的 16.0 盎司的净容量. 做了一个实验, 从机器的产品中各取一个随机样本.

机器 1	16.03	16.04	16.05	16.05	16.02	16.01	15.96	15.98	16.02	15.99
机器 2	16.02	15.97	15.96	16.01	15.99	16.03	16.04	16.02	16.01	16.00

(a) 叙述你认为在该实验中应检验的假设.

(b) 用 α=0.05 检验这些假设, 你的结论是什么?

(c) 确定本次检验的 P 值.

(d) 确定两台机器的平均充装容量差的一个 95%置信区间.

2.10 两类塑料都适用于一家电子计算器厂. 塑料的抗断强度是重要的. 已知 $\sigma_1 = \sigma_2 = 1.0$ psi. 从 $n_1 = 10$ 和 $n_2 = 12$ 的随机样本得 $\overline{y}_1 = 162.5$ 和 $\overline{y}_2 = 155.0$. 公司不采用塑料 1, 除非它的抗断强度超过塑料 2 的抗断强度至少 10 psi. 根据样本信息, 他们应该用塑料 1 吗? 为回答这个问题, 建

立合适的假设并用 $\alpha=0.01$ 进行检验, 构造一个关于抗断强度均值差的 99%置信区间.

*2.11 下面是两种不同配方的化学照明弹的燃烧时间 (单位: 分钟), 设计工程师感兴趣于两者燃烧时间的均值和方差.

型 1	65	81	57	66	82	82	67	59	75	70
型 2	64	71	83	59	65	56	69	74	82	79

(a) 用 $\alpha = 0.05$ 检验两个方差相等的假设.

(b) 用 (a) 的结果, 检验平均燃烧时间相等的假设. 用 $\alpha = 0.05$, 检验的 P 值是多少?

(c) 讨论正态性假定在本问题中的作用. 检验两种类型的照明弹的正态性假设.

*2.12 G. Z. Yin 与 D. W. Jillie 在 *Solid State Technology*(May, 1987) 上发表的一篇论文 "Orthogonal Design for Process Optimization and Its Application to Plasma Etching" 描述了一个实验, 在集成电路生产中, 要确定 C_2F_6 流率对蚀刻硅片的均匀度是否有影响. 所有的试验顺序都是随机的. 两种流率的数据如下:

C_2F_6 流 (SCCM)	均匀度观测值					
	1	2	3	4	5	6
125	2.7	4.6	2.6	3.0	3.2	3.8
200	4.6	3.4	2.9	3.5	4.1	5.1

(a) C_2F_6 流率对蚀刻均匀度的均值有无影响? 用 $\alpha = 0.05$.

(b) (a) 中检验的 P 值是多少?

(c) C_2F_6 流率在蚀刻均匀度方面对片与片的变异性有无影响? 用 $\alpha = 0.05$.

(d) 用实验数据画盒图支持你的阐述.

2.13 在一个化工装置中安装了一台新的过滤器. 在安装之前, 一个随机样本提供了关于杂质百分率的信息如下: $\bar{y}_1 = 12.5$, $S_1^2 = 101.17$, $n_1 = 8$. 安装之后, 一个随机样本提供 $\bar{y}_2 = 10.2$, $S_2^2 = 94.73$, $n_2 = 9$.

(a) 你能得出两个方差相等的结论吗? 用 $\alpha = 0.05$.

(b) 过滤器是否显著地减少了杂质的百分比? 用 $\alpha = 0.05$.

2.14 光致抗蚀剂是一种应用在半导体晶片上的光敏材料. 通过它, 电路可以刻制在晶片上. 用后, 烘焙涂层的晶片以除去光致抗蚀剂混合物的溶媒以硬化保护层. 下表列出了 8 个晶片在两种不同温度下烘焙的光致抗蚀剂厚度 (kA). 假设所有的炉次都是随机的.

95 ℃	11.176	7.089	8.097	11.739	11.291	10.759	6.467	8.315
100 ℃	5.263	6.748	7.461	7.015	8.133	7.418	3.772	8.963

(a) 有无证据认为较高的烘焙温度导致晶片的光致抗蚀剂厚度的均值较低? 用 $\alpha = 0.05$.

(b) (a) 中进行的检验的 P 值是多少?

(c) 确定均值差异的一个 95%置信区间. 提供一个该区间的实际解释.

(d) 画点图来说明这次实验的结论.

(e) 检验光致抗蚀剂厚度的正态性假设.

(f) 确定用来探测均值差异为 2.5 kA 的检验的势.

(g) 计算探测均值差异为 1.5 kA、势至少为 0.9 所需的样本量.

*2.15 手机的外壳是用注塑模工艺生产的. 部件被取出之前在模具中冷却的时间会影响手机的外观缺陷的发生. 加工后, 目测检查, 给出它们的外观 1 到 10 的得分, 10 对应完美的部件, 1 对应最差的部件. 用两种冷却时间 (10 秒和 20 秒) 做实验, 在每种冷却时间水平上评价 20 个外壳. 这个实验的所有 40 个观测都是按照随机次序进行的. 数据如下.

10 秒	1	2	1	3	5	1	5	2	3	5
	3	6	5	3	2	1	6	8	2	3
20 秒	7	8	5	9	5	8	6	4	6	7
	6	9	5	7	4	6	8	5	8	7

(a) 是否有证据认为更长的冷却时间会导致更少的外观缺陷? 用 $\alpha = 0.05$.

(b) (a) 中进行的检验的 P 值是多少?

(c) 确定均值差异的一个 95%置信区间. 提供一个该区间的实际解释.

(d) 画点图来解释这次实验的结论.

(e) 检验实验数据的正态性假设.

2.16 在对等离子体蚀刻器的合格性实验中取得了在硅片上蚀刻均匀度的 20 个观测值. 这些数据如下:

5.34	6.65	4.76	5.98	7.25	6.00	7.55	5.54	5.62	6.21
5.97	7.35	5.44	4.39	4.98	5.25	6.35	4.61	6.00	5.32

(a) 构造估计 σ^2 的 95%置信区间.

(b) 用 $\alpha = 0.05$ 检验假设 $\sigma^2 = 1.0$. 你的结论是什么?

(c) 讨论正态性假设及其在本题中的作用.

(d) 构造正态概率图检验正态性. 你的结论是什么?

2.17 由 12 位测量员测定滚珠轴承的直径, 每人用两种不同的测径器, 其结果是

测量员	测径器 1	测径器 2	测量员	测径器 1	测径器 2
1	0.265	0.264	7	0.267	0.264
2	0.265	0.265	8	0.267	0.265
3	0.266	0.264	9	0.265	0.265
4	0.267	0.266	10	0.268	0.267
5	0.267	0.267	11	0.268	0.268
6	0.265	0.268	12	0.265	0.269

(a) 用 α=0.05 检验两个被选样本测量总体的均值有无显著性差异?

(b) (a) 中进行的检验的 P 值是多少?

(c) 构造两种类型测径器测量的直径均值差异的一个 95%置信区间.

*2.18 *Journal of Strain Analysis*(Vol.18, no.2, 1983) 上的一篇论文比较了几种预测钢板梁的抗剪强度的方法. 对其中的两种方法 (即 Karlsruhe 法和 Lehigh 法), 9 根梁的数据 (预测荷载与观测荷载的比值) 如下:

梁	S1/1	S2/1	S3/1	S4/1	S5/1	S2/1	S2/2	S2/3	S2/4
Karlsruhe 法	1.186	1.151	1.322	1.339	1.200	1.402	1.365	1.537	1.559
Lehigh 法	1.061	0.992	1.063	1.062	1.065	1.178	1.037	1.086	1.052

(a) 是否有证据认为两种方法得到的均值存在差异? 用 $\alpha = 0.05$.

(b) (a) 中进行的检验的 P 值是多少?

(c) 构造预测荷载与观测荷载两种均值差异的一个 95%置信区间.

(d) 研究两种样本的正态性假设.

(e) 研究两种方法的比的差异的正态性假设.

(f) 讨论配对 t 检验中的正态性假设的作用.

2.19 研究 ABS 塑胶管的两种不同配方的负荷变形温度. 用两种配方各进行 12 次试验, 负荷变形温度 (单位: °F) 结果如下:

配方 1			配方 2		
206	193	192	177	176	198
188	207	210	197	185	188
205	185	194	206	200	189
187	189	178	201	197	203

(a) 分别构造两个样本的正态概率图. 该图是否支持正态性假定和两样本方差相等的论点?

(b) 数据是否支持配方 1 的负荷变形温度超过配方 2 的论点? 用 $\alpha = 0.05$.

(c) (b) 中进行检验的 P 值是多少?

2.20 再研究思考题 2.19 的数据. 数据是否支持配方 1 的负荷变形温度超过配方 2 至少 3°F 的论点?

*2.21 半导体生产中, 湿化学蚀刻通常用于金属化之前去除晶片基座的硅元素. 蚀刻率是这道工序的重要特性. 要评估两种不同的蚀刻方法. 用每种方法蚀刻随机选择的 8 个晶片, 观测的蚀刻率如下 (单位: mils/min).

方法 1		方法 2	
9.9	10.6	10.2	10.6
9.4	10.3	10.0	10.2
10.0	9.3	10.7	10.4
10.3	9.8	10.5	10.3

(a) 数据是否表明所有方法有相同的蚀刻率? 用 $\alpha = 0.05$ 并假定等方差.

(b) 确定蚀刻率均值差异的一个 95%置信区间.

(c) 用正态概率图研究正态性假设和方差相等的适当性.

2.22 根据人体的吸收速度比较两种流行的止痛药. 有人说药片 1 的吸收度是药片 2 的两倍. 设 σ_1^2 与 σ_2^2 已知. 导出

$$H_0: 2\mu_1 = \mu_2, \qquad H_1: 2\mu_1 \neq \mu_2$$

的检验统计量.

2.23 设要检验

$$H_0: \mu_1 = \mu_2, \qquad H_1: \mu_1 \neq \mu_2$$

其中 σ_1^2 与 σ_2^2 已知. 因抽样资源所限, 给定 $n_1 + n_2 = N$. 为了得到最大功效的检验, 应该如何在两个总体之间分配 N 个观测值.

2.24 推导 (2.46) 式, 得到正态分布方差的一个 $100(1-\alpha)$%置信区间.

2.25 推导 (2.50) 式, 得到比值 σ_1^2/σ_2^2 的一个 $100(1-\alpha)$%置信区间, 其中 σ_1^2 与 σ_2^2 是两个正态分布的方差.

2.26 推导一个公式, 以得到 $\sigma_1^2 \neq \sigma_2^2$ 的两个正态分布均值差的一个 $100(1-\alpha)$%置信区间. 将你的公式用于硅酸盐水泥实验数据, 确定一个 95%的置信区间.

2.27 构造一个数据集, 使得配对 t 检验统计量非常大, 但通常的两样本或合并 t 检验统计量却较小. 描述你怎样创建数据. 这能否给你一些如何运用配对 t 检验的直观认识?

*2.28 考虑思考题 2.11 描述的实验. 如果两个照明弹的燃烧时间差为 2 分钟, 确定检验的势. 计算势至少为 0.9, 检测燃烧时间均值差异在 1 分钟所需要的样本量.

2.29 再考虑思考题 2.9 描述的瓶子充装问题. 假定两总体方差未知但相等, 重做这个问题.

2.30 考虑思考题 2.9 的数据, 如果两种机器充装量均值的差为 0.25 盎司, 用于思考题 2.9 的检验的势的大小是多少? 如果充装量均值的真实差为 0.25 盎司, 势至少为 0.9 的样本量是多大?

第 3 章　单因子实验：方差分析

本章纲要

3.1　一个例子
3.2　方差分析
3.3　固定效应模型的分析
3.3.1　总平方和的分解
3.3.2　统计分析
3.3.3　模型参数的估计
3.3.4　不平衡数据
3.4　模型适合性检测
3.4.1　正态性假设
3.4.2　依时间序列的残差图
3.4.3　残差与拟合值的关系图
3.4.4　残差与其他变量的关系图
3.5　结果的实际解释
3.5.1　回归模型
3.5.2　处理均值的比较
3.5.3　均值的图解比较法
3.5.4　对照法
3.5.5　正交对照法
3.5.6　用来比较全部对照的 Scheffé方法
3.5.7　处理均值的配对比较法
3.5.8　将各个处理均值与一个控制进行比较
3.6　计算机输出示例
3.7　确定样本量
3.7.1　抽检特性曲线
3.7.2　规定标准差的增量
3.7.3　置信区间的估计方法
3.8　寻找分散效应
3.9　方差分析的回归处理法
3.9.1　模型参数的最小二乘估计
3.9.2　一般回归显著性检验
3.10　方差分析中的非参数方法
3.10.1　Kruskal-Wallis 检验法
3.10.2　关于秩变换的一般评论
第 3 章补充材料
S3.1　因子效应的定义
S3.2　期望均方
S3.3　σ^2 的置信区间
S3.4　处理均值的联合置信区间
S3.5　定量因子的回归模型
S3.6　关于估计函数的更多内容
S3.7　回归与方差分析的关系

第 2 章讨论了比较两个条件或两种处理的方法. 例如, 硅酸盐水泥的粘合抗拉强度实验涉及了两种不同砂浆配方. 描述这类实验的另一种方法是二水平的单因子实验, 这里的因子是指砂浆配方, 二水平就是两种不同的配方. 很多这类实验的因子都是多于两个水平的. 本章侧重介绍关于具有 a 个水平 (或 a 个处理) 的单因子实验的设计与分析方法. 假定实验都是完全随机化的.

3.1　一 个 例 子

在集成电路的许多生产步骤中, 晶片被一层材料 (如二氧化硅或某种金属) 完全覆盖. 通过对掩模的蚀刻有选择性地除去不需要的材料, 从而创建电路模板、电互连以及必须扩散的或者金属沉积的区域. 等离子蚀刻工序在这个操作中被广泛使用, 特别是在几何对象比较小的情况下的

应用. 图 3.1 展示了一种典型的单晶片蚀刻设备的重要特征. 射频 (RF) 发生器提供能源, 使得电极之间的间隙产生等离子. 等离子体的化学种类是由所使用的特定气体决定的. 碳氟化合物, 比如 CF_4(四氟甲烷) 或 C_2F_6(六氟乙烷), 通常被用在等离子蚀刻上. 但是根据应用情况的不同, 也常使用其他的气体或混合气体.

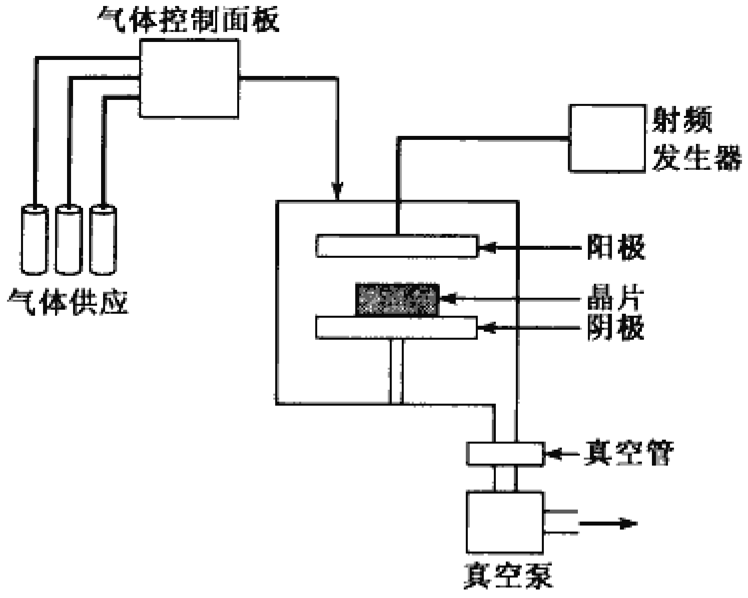


图 3.1 单晶片等离子蚀刻设备

工程师要研究这套设备的 RF 功率设置与蚀刻率间的关系. 实验目的是建立蚀刻率与 RF 功率之间关系的模型, 以确定达到所需的目标蚀刻率的功率设置. 她选定了气体 (C_2F_6) 和间隙 (0.80 cm), 想检验 RF 功率的 4 个水平: 160W, 180W, 200W 和 220W. 她决定在 RF 功率的每个水平上检验 5 个晶片.

这是一个因子**水平** $a=4$ 和**重复** $n=5$ 的单因子实验. 这 20 个试验都是按照随机次序进行的. 一种生成随机次序的更高效方法是将 20 个试验输入电子表格 (Excel), 用 Rand() 函数产生一列随机数, 以此排序.

假设从这个过程中获得的试验次序是

试验次序	Excel 随机数 (已排序)	功率	试验次序	Excel 随机数 (已排序)	功率
1	12 417	200	11	49 813	220
2	18 369	220	12	52 286	220
3	21 238	220	13	57 102	160
4	24 621	160	14	63 548	160
5	29 337	160	15	67 710	220
6	32 318	180	16	71 834	180
7	36 481	200	17	77 216	180
8	40 062	160	18	84 675	180
9	43 289	180	19	89 323	200
10	49 271	200	20	94 037	200

为防止未知讨厌变量的影响, 随机化试验次序是必要的. 因为实验中讨厌变量的变化也许会超出控制范围, 从而损害实验结果. 为了说明这一点, 假定我们按原始的非随机化次序做 20 个晶片试验 (也就是, 前 5 个功率用 160W, 接下来的 5 个功率用 180W, 依次类推). 如果蚀刻设

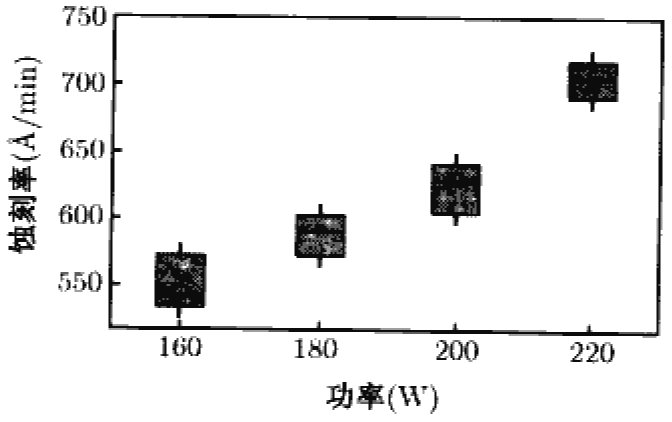
备有热身效应, 则运行越长, 观测的蚀刻率读数就越低, 热身效应将潜在地损害数据, 从而破坏实验的有效性.

假定工程师按照我们确定的随机次序进行试验. 得到的蚀刻率的观测值如表 3.1 所示.

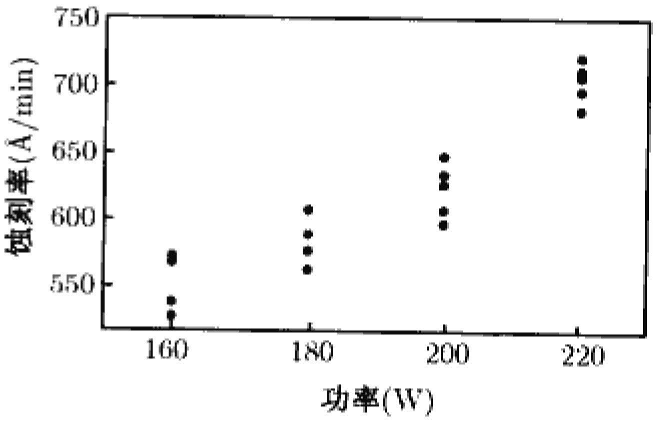
表 3.1　晶片蚀刻实验中的蚀刻率数据　(单位: Å/ min)

RF 功率 (W)	观测值					总和	平均值
	1	2	3	4	5		
160	575	542	530	539	570	2 756	551.2
180	565	593	590	579	610	2 937	587.4
200	600	651	610	637	629	3 127	625.4
220	725	700	715	685	710	3 535	707.0

用**图示方法**来检查实验数据的想法也很好. 图 3.2a(**盒图**) 展示了 RF 功率的每个水平的蚀刻率, 图 3.2b 是蚀刻率关于 RF 功率的**散点图**. 两类图都表明蚀刻率随着功率设置的增加而增加. 但没有明显证据表明蚀刻率相对于均值的变异性依赖于功率设置. 基于这个简单的图形分析, 我们猜想: (1) RF 功率设置影响蚀刻率; (2) 功率设置提高时, 蚀刻率增加.



(a) 比较的盒图



(b) 散点图

图 3.2　蚀刻率数据的盒图和散点图

假定想更客观地分析这些数据. 特别地, 假定希望检验在 RF 功率的所有 a=4 水平下蚀刻率均值间的差异. 于是, 我们想要检验所有的 4 个均值是否都相等. 粗看起来, 这个问题可以用所有 6 个可能的配对比较的 t 检验来解决. 但是, 这并不是解决这个问题的最好方法. 首先, 做 6 个 t 检验效率很低, 事倍功半. 第二, 这些配对比较的执行增大了第一类误差. 假定 4 个均值都相等, 如果我们选择 $\alpha = 0.05$, 则在每一单个比较上作出正确决策的概率是 0.95. 但是, 所有 6 个比较都是正确决策的概率就远远小于 0.95, 所以, 第一类错误被增大了.

方差分析是检验若干个均值相等的比较好的方法. 不过, 方差分析的用途比解决上述问题要更为广泛得多. 它可能是在统计推断领域中最有用的方法.

3.2　方 差 分 析

设想我们有 a 个处理或想要比较的某一**单因子**的不同**水平**. a 个处理的每一个中的观测响应都是一个随机变量. 数据的表示方式可如表 3.2 那样. 表 3.2 中的任一值 (例如, y_{ij}) 代表第 i 种因子水平 (处理) 下所取的第 j 个观测值. 一般说来, 在第 i 种处理下有 n 个观测值. 表 3.2

是表 3.1 所示晶片蚀刻实验数据的一般情形.

表 3.2 单因子实验的典型数据

处理 (水平)	观测值				总和	平均值
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1\cdot}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a.}$ $y_{\cdot\cdot}$	$\bar{y}_{a.}$ $\bar{y}_{\cdot\cdot}$

1. **数据的模型**

我们发现用模型来描述这些观测值是有用的, 一种方法是将模型写成

$$y_{ij} = \mu_i + \varepsilon_{ij} \qquad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, n \end{cases} \tag{3.1}$$

其中 y_{ij} 是第 ij 个观测值, μ_i 是第 i 个因子水平 (或处理) 的均值, ε_{ij} 是**随机误差**分量, 它合并了实验中其他的变异来源, 包括测量、来自不可控因子的变异、用于处理的实验单元间的差异 (像实验材料, 等等) 以及过程中广泛的背景噪声因子 (像时间的变异性, 环境变量的效应, 等等). 为方便起见, 假定误差的均值为零, 从而 $\mathrm{E}(y_{ij}) = \mu_i$.

(3.1) 式被称为**均值模型**(means model). 可以换一种方式写出数据的模型, 定义

$$\mu_i = \mu + \tau_i, \quad i = 1, 2, \cdots, a$$

则 (3.1) 式变为

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, n \end{cases} \tag{3.2}$$

在模型的这种形式中, μ 是所有处理的共同参数, 叫做**总均值**(overall mean), τ_i 是唯一对应于第 i 个处理的参数, 叫做第 i 个**处理效应**(treatment effect), (3.2) 式通常被称为**效应模型**(effects model).

均值模型和效应模型两者都是**线性统计模型**(linear statistical model), 即响应变量 y_{ij} 是模型参数的线性函数. 尽管模型的这两种形式都是有用的, 但在实验设计文献中效应模型更常见. 效应模型中, μ 是一个常量, 而处理效应 τ_i 表示应用具体处理时与常量 μ 的偏差, 这使得该模型更具直观性.

因为只研究一个因子, 所以 (3.2) 式 [或 (3.1) 式] 也叫做**一种方式的方差分析或单因子方差分析 (ANOVA)**模型. 此外, 我们要求实验是以随机顺序进行的, 以便用于处理的环境 (通常叫做**实验单元**) 尽可能一致. 于是, 这一实验的设计属于**完全随机化设计**. 我们的目标是检验关于处理均值的假设并估计处理均值. 在假设检验中, 假定模型误差是独立的正态分布随机变量, 其均值为零, 方差为 σ^2. 假定方差 σ^2 在该因子的各个水平下都是常量. 因此, 观测值

$$y_{ij} \sim N(\mu + \tau_i, \sigma^2)$$

而且这些观测值是相互独立的.

2. 固定因子还是随机因子?

统计模型 (3.2) 式描述了关于处理效应的两种不同情况. 第一种情况是, a 个处理可以由实验者具体选定. 此时我们想检验关于处理均值的假设, 因此所得结论仅适用于该分析中所考虑的因子水平. 这些结论不能推广到未曾明确考虑的相似处理中去. 我们可能还想去估计模型参数 (μ, τ_i, σ^2). 这叫做**固定效应模型**(fixed effects model). 另一种情况是, a 个处理可以看作是来自一个较大的处理总体的一个**随机样本**(random sample). 在这种情况下, 我们希望能够把这些结论 (根据处理样本所得的) 推广到总体的所有处理中去, 而不管它们在分析中是否被明确考虑. 这里, τ_i 是随机变量, 而关于个别变量的知识相对说来是无用的. 因而, 我们要检验关于 τ_i 的变异性假设并试图估计这一变异性. 这叫做**随机效应模型**或**方差分量模型**. 第 13 章将进一步讨论含随机因子的实验.

3.3　固定效应模型的分析

本节研究固定效应模型的单因子方差分析. 令 $y_{i.}$ 表示第 i 个处理的观测值的总和, $\overline{y}_{i.}$ 表示第 i 个处理的观测值的平均值. 类似地, 令 $y_{..}$ 表示全部观测值的总和, $\overline{y}_{..}$ 表示全部观测值的总平均值. 用符号表示即

$$\begin{aligned} y_{i.} &= \sum_{j=1}^{n} y_{ij}, \quad \overline{y}_{i.} = y_{i.}/nm, \qquad i = 1, 2, \cdots, a \\ y_{..} &= \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}, \quad \overline{y}_{..} = y_{..}/N \end{aligned} \tag{3.3}$$

其中 $N{=}an$ 是观测值的总个数. 可见, 下标 "点" 号意味着对相应的下标求和.

我们的兴趣在于检验 a 个处理均值的等式, 即

$$\mathrm{E}(y_{ij}) = \mu + \tau_i = \mu_i,\ i = 1, 2, \cdots, a.$$

这里较合理的假设是

$$\begin{aligned} &H_0:\ \mu_1 = \mu_2 = \cdots = \mu_a \\ &H_1:\ \mu_i \neq \mu_j, \quad \text{至少对于一对 } (i, j) \end{aligned} \tag{3.4}$$

在效应模型中, 我们将第 i 个处理均值 μ_i 分成两个分量, 即 $\mu_i = \mu + \tau_i$. 我们通常认为 μ 是总平均, 所以,

$$\frac{\sum\limits_{i=1}^{a} \mu_i}{a} = \mu$$

这个定义意味着

$$\sum_{i=1}^{a} \tau_i = 0$$

即处理或因子效应可以被认为是与总均值的偏差①. 因此, 根据处理效应 τ_i, 前面的假设的一个等价写法就是

① 这个论题的更多信息见第 3 章的补充材料.

$$H_0:\ \tau_1=\tau_2=\cdots=\tau_a=0$$

$$H_1:\ \tau_i\neq 0 \qquad \text{至少对于一个 } i$$

从而, 我们说检验处理均值相等或检验处理效应 (τ_i) 为零. 检验 a 个处理均值相等的合适方法就是方差分析.

3.3.1 总平方和的分解

方差分析(analysis of variance) 这一名称来源于把总变异性分解为它的分量. 总校正平方和

$$SS_T=\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{..})^2$$

是数据总变异性的一个度量. 直观看来, 这是合理的, 因为如果将 SS_T 除以一个适当的自由度 (在现在的情况下, 是 $an-1=N-1$), 就得到 y 的**样本方差**. 当然, 样本方差是变异性的标准度量.

总校正平方和 SS_T 可以写为

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{..})^2=\sum_{i=1}^{a}\sum_{j=1}^{n}[(\overline{y}_{i.}-\overline{y}_{..})+(y_{ij}-\overline{y}_{i.})]^2 \tag{3.5}$$

即

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{..})^2=n\sum_{i=1}^{a}(\overline{y}_{i.}-\overline{y}_{..})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{i.})^2+2\sum_{i=1}^{a}\sum_{j=1}^{n}(\overline{y}_{i.}-\overline{y}_{..})(y_{ij}-\overline{y}_{i.})$$

但是, 等式最后的交叉乘积项为零, 因为

$$\sum_{j=1}^{n}(y_{ij}-\overline{y}_{i.})=y_{i.}-n\overline{y}_{i.}=y_{i.}-n(y_{i.}/n)=0$$

因此, 我们有

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{..})^2=n\sum_{i=1}^{a}(\overline{y}_{i.}-\overline{y}_{..})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{i.})^2 \tag{3.6}$$

(3.6) 式就是方差分析的公式, 它表明, 用总校正平方和来度量的一组数据的总变异性可以分解为: 处理平均值与总平均值之差的平方和, 再加上在处理内部的观测值与处理平均值之差的平方和. 现在, 观测得的处理平均值与总平均值之间的差就是处理均值之间的差的度量, 而处理内部的观测值与处理平均值之差却只是随机误差. 于是, 可以将 (3.6) 式用符号写为

$$SS_T=SS_{\text{处理}}+SS_E$$

其中 $SS_{\text{处理}}$ 叫做处理 (即, 在处理之间的) 平方和, 而 SS_E 叫做误差 (即, 处理内部的) 平方和. 因为有 $an=N$ 个总观测值, 所以, SS_T 有 $N-1$ 个自由度. 又因为有 a 个因子水平 (以及 a 个处理平均值), 所以, $SS_{\text{处理}}$ 有 $a-1$ 个自由度. 最后, 任一处理内部有 n 个重复, 给出 $n-1$ 个自由度, 它可以用于估计实验的误差, 因为有 a 个处理, 从而误差有 $a(n-1)=an-a=N-a$ 个自由度.

深入剖析基本方差分析恒等式的右边两项是有益的. 考虑误差平方和

$$SS_E=\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{i.})^2=\sum_{i=1}^{a}\left[\sum_{j=1}^{n}(y_{ij}-\overline{y}_{i.})^2\right]$$

由这一形式容易看出方括号内的项除以 $n-1$ 就是第 i 个处理的样本方差, 或

$$S_i^2 = \frac{\sum_{j=1}^{n}(y_{ij}-\overline{y}_{i\cdot})^2}{n-1} \qquad i=1,2,\cdots,a$$

现在, a 个样本方差组合起来给出公共总体方差的一个估计如下:

$$\frac{(n-1)S_1^2+(n-1)S_2^2+\cdots+(n-1)S_a^2}{(n-1)+(n-1)+\cdots+(n-1)} = \frac{\sum_{i=1}^{a}\left[\sum_{j=1}^{n}(y_{ij}-\overline{y}_{i\cdot})^2\right]}{\sum_{i=1}^{a}(n-1)} = \frac{SS_E}{(N-a)}$$

于是, $SS_E/(N-a)$ 就是在 a 个处理中每一个处理公共方差的一个综合估计.

类似地, 如果 a 个处理均值之间没有差别, 就可以用处理平均值对总平均值的偏差平方来估计 σ^2. 特别地,

$$\frac{SS_{\text{处理}}}{a-1} = \frac{n\sum_{i=1}^{a}(\overline{y}_{i\cdot}-\overline{y}_{\cdot\cdot})^2}{a-1}$$

是处理均值都相等时 σ^2 的一个估计. 这一点的理由可以很直观地看出: 量 $\sum_{i=1}^{a}(\overline{y}_{i.}-\overline{y}_{..})^2/(a-1)$ 估计处理平均值的方差 σ^2/n, 所以当处理均值之间没有差别时, 应该用 $n\sum_{i=1}^{a}(\overline{y}_{i.}-\overline{y}_{..})^2/(a-1)$ 来估计 σ^2.

方差分析恒等式 [(3.6) 式] 给 σ^2 提供了两个估计 —— 一个根据处理内部的固有变异性, 一个根据处理之间的变异性. 如果处理均值没有差别, 这两种估计就应该十分相近; 否则, 我们就推测观测到的差异是由于处理均值之间的差异所引起的. 尽管我们已经用直观论证研究了这一结果, 但也可运用更形式些的推导.

量 $MS_{\text{处理}} = \frac{SS_{\text{处理}}}{a-1}$ 与 $MS_E = \frac{SS_E}{N-a}$ 称为**均方**. 现在考察这些均方的**期望值**. 考虑

$$\begin{aligned}
\mathrm{E}(MS_E) &= \mathrm{E}\left(\frac{SS_E}{N-a}\right) = \frac{1}{N-a}\mathrm{E}\left[\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\overline{y}_{i.})^2\right] \\
&= \frac{1}{N-a}\mathrm{E}\left[\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}^2-2y_{ij}\overline{y}_{i.}+\overline{y}_{i.}^2)\right] \\
&= \frac{1}{N-a}\mathrm{E}\left[\sum_{i=1}^{a}\sum_{j=1}^{n}y_{ij}^2-2n\sum_{i=1}^{a}\overline{y}_{i.}^2+n\sum_{i=1}^{a}\overline{y}_{i.}^2\right] \\
&= \frac{1}{N-a}\mathrm{E}\left[\sum_{i=1}^{a}\sum_{j=1}^{n}y_{ij}^2-\frac{1}{n}\sum_{i=1}^{a}y_{i.}^2\right]
\end{aligned}$$

将模型 [(3.1) 式] 代入此式, 得

$$\mathrm{E}(MS_E) = \frac{1}{N-a}\mathrm{E}\left[\sum_{i=1}^{a}\sum_{j=1}^{n}(\mu+\tau_i+\varepsilon_{ij})^2-\frac{1}{n}\sum_{i=1}^{a}\left(\sum_{j=1}^{n}(\mu+\tau_i+\varepsilon_{ij})\right)^2\right]$$

现在因为 $E(\varepsilon_{ij})=0$, 所以在进行平方运算并对方括号内的量取期望值时, 涉及 ε_{ij}^2 和 $\varepsilon_{i.}^2$ 的项都可以分别用 σ^2 与 $n\sigma^2$ 来代替. 而且, 所有含有 ε_{ij} 的交叉乘积的期望都是零. 因此, 在进行了平方运算和取期望之后, 最后的等式变为

$$\mathrm{E}(MS_E)=\frac{1}{N-a}\left[N\mu^2+n\sum_{i=1}^{a}\tau_i^2+N\sigma^2-N\mu^2-n\sum_{i=1}^{a}\tau_i^2-a\sigma^2\right]$$

即
$$\mathrm{E}(MS_E)=\sigma^2$$

用类似的方法, 我们亦可证明[①]

$$\mathrm{E}(MS_{处理})=\sigma^2+\frac{n\sum\limits_{i=1}^{a}\tau_i^2}{a-1}$$

正如我们启发性的讨论, $MS_E=SS_E/(N-a)$ 可用于估计 σ^2, 而且如果在处理均值间没有差异时 (意味着 $\tau_i=0$), $MS_{处理}=SS_{处理}/(a-1)$ 亦可估计 σ^2. 不过, 注意如果处理均值确实不同, 则处理均方的期望值大于 σ^2.

这清楚地表明, 检验处理均值之间没有差异这一假设可以代之以比较 $MS_{处理}$ 和 MS_E 来实行. 我们现在就考虑这一比较如何进行.

3.3.2 统计分析

我们现在来研究检验处理均值之间没有差异这一假设 ($H_0:\mu_1=\mu_2=\cdots=\mu_a$ 或者 $H_0:\tau_1=\tau_2=\cdots=\tau_a=0$) 如何进行. 因为我们已经假定误差 ε_{ij} 是独立服从正态分布的, 其均值为零, 方差是 σ^2, 所以观测值 y_{ij} 也是独立服从正态分布的, 其均值为 $\mu+\tau_i$, 方差为 σ^2. 于是, SS_T 是正态随机变量的平方和, 因此, 可以证明 SS_T/σ^2 服从自由度为 $N-1$ 的卡方分布. 进一步, 当零假设 $H_0:\tau_i=0$ 为真时, SS_E/σ^2 服从自由度为 $N-a$ 的卡方分布, 并且 $SS_{处理}/\sigma^2$ 服从自由度为 $a-1$ 的卡方分布. 然而, 因为 $SS_{处理}$ 与 SS_E 加起来等于 SS_T, 所以这 3 个平方和并不一定是独立的. 下面的定理用来建立 SS_E 与 $SS_{处理}$ 的独立性, 它是 William G. Cochran 定理的一种特殊形式.

定理 3.1 Cochran 定理

设 Z_i 是 NID(0, 1), $i=1,2,\cdots,v$, $\sum\limits_{i=1}^{\nu}Z_i^2=Q_1+Q_2+\cdots+Q_s$, 其中 $s\leqslant v$, Q_i 的自由度为 $v_i(i=1,2,\cdots,s)$, 则当且仅当 $\nu=\nu_1+\nu_2+\cdots+\nu_s$ 时, $Q_1,Q_2,\cdots,Q_s$ 是相互独立的卡方分布随机变量, 其自由度分别为 $v_1,v_2,\cdots,v_s$.

因为 $SS_{处理}$ 和 SS_E 的自由度加起来等于总的自由度 $N-1$, 所以, Cochran 定理蕴含了 $SS_{处理}/\sigma^2$ 与 SS_E/σ^2 是独立分布的 χ^2 随机变量. 因此, 当处理均值之间没有差异的零假设为真时, 比值

$$F_0=\frac{SS_{处理}/(a-1)}{SS_E/(N-a)}=\frac{MS_{处理}}{MS_E} \tag{3.7}$$

服从自由度为 $a-1$ 与 $N-a$ 的 F 分布. (3.7) 式就是关于处理均值之间没有差异的假设的**检验统计量**.

---

① 见第 3 章补充材料.

由均方的期望可见, MS_E 是 σ^2 的一个无偏估计. 而且在零假设下, $MS_{处理}$ 也是 σ^2 的一个无偏估计. 然而, 如果零假设不真, 则 $MS_{处理}$ 的期望值大于 σ^2. 因此, 在备择假设下, 检验统计量 [(3.7) 式] 的分子的期望值大于分母的期望值, 所以, 当检验统计量的值太大时, 我们应该拒绝 H_0. 这意味着是一个上尾部的, 单尾拒绝域. 因此, 当

$$F_0 > F_{\alpha,a-1,N-a}$$

时, 应该拒绝 H_0, 其中 F_0 由 (3.7) 式算得, 并且得出在不同的处理方法之间存在着差异的结论. 另外, 我们也可以利用 P 值方法进行决策.

平方和的计算有几种方法. 一种直接的方法是利用定义:

$$y_{ij} - \bar{y}_{..} = (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.}),$$

用一个展开式计算每个观测的这三项, 然后, 将平方相加得到 $SS_{处理}$、SS_T 和 SS_E. 另一种方法是改写并简化 (3.6) 式中的 $SS_{处理}$ 和 SS_T 的定义, 有

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{3.8}$$

$$SS_{处理} = \frac{1}{n}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N} \tag{3.9}$$

$$SS_E = SS_T - SS_{处理} \tag{3.10}$$

这种方法很好, 因为一些计算器被设计成把输入数的和存储在一个存储器, 而输入数的平方和存储在另一个存储器. 因此每一个数据只需输入一次. 实际上, 我们用计算机软件进行这一操作.

表 3.3 总结了检验过程, 我们称之为**方差分析表**, 即 ANOVA 表.

表 3.3 单因子方差分析表, 固定效应模型

方差来源	平方和	自由度	均方	F_0
处理间	$SS_{处理} = n\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	$a-1$	$MS_{处理}$	$F_0 = \frac{MS_{处理}}{MS_E}$
误差 (处理内)	$SS_E = SS_T - SS_{处理}$	$N-a$	MS_E	
总和	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot\cdot})^2$	$N-1$		

例 3.1 晶片蚀刻实验

为说明方差分析方法, 我们回到 3.1 节讨论过的例子. 开发工程师要确定 RF 功率设置是否影响蚀刻率, 她进行了 RF 功率有 4 个水平和 5 次重复的完全随机化实验. 为方便起见, 此处重复表 3.1 的数据如下.

RF 功率 (W)	蚀刻率的观测 (Å/ min)					总和	平均值
	1	2	3	4	5	y_i	$\bar{y}_{i\cdot}$
160	575	542	530	539	570	2 756	551.2
180	565	593	590	579	610	2 937	587.4
200	600	651	610	637	629	3 127	625.4
220	725	700	715	685	710	3 535	707.0
						$y_{..}$=12 355	$\bar{y}_{..}$ = 617.75

我们用方差分析检验原假设 $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ 与备择假设 H_1: 部分均值不等. 用 (3.8) 式、(3.9) 式和 (3.10) 式计算所需的平方和如下:

$$SS_T = \sum_{i=1}^{4}\sum_{j=1}^{5} y_{ij}^2 - \frac{y_{..}^2}{N} = (575)^2 + (542)^2 + \cdots + (710)^2 - (12\ 355)^2/20 = 72\ 209.75$$

$$SS_{处理} = \frac{1}{n}\sum_{i=1}^{4} y_{i.}^2 - \frac{y_{..}^2}{N} = \frac{1}{5}[(2\ 756)^2 + \cdots + (3\ 535)^2] - (12\ 355)^2/20 = 66\ 870.55$$

$$SS_E = SS_T - SS_{处理} = 72\ 209.75 - 66\ 870.55 = 5\ 339.20$$

通常, 这些计算是通过使用具有分析从已设计的实验中获得的数据的能力的软件包在计算机上进行的.

方差分析概括在表 3.4 中. RF 功率或处理间的均方值 (22 290.18) 是处理内部或误差均方值 (333.70) 的好几倍. 这表明处理均值都相等是不大可能的. 形式上, 计算 F 的比值 F_0=22 290.18/333.70=66.80, 并与分布 $F_{3,16}$ 合适的上尾百分点比较. 假定实验者选择 $\alpha = 0.05$, 从附表VI我们发现 $F_{0.05,3,16} = 3.24$, 因为 $F_0 = 66.80 > 3.24$, 所以我们拒绝 H_0, 并得出处理均值不相同的结论. 也就是说, RF 功率的设置对平均蚀刻率有显著影响. 我们还可以计算检验统计量的 P 值. 图 3.3 给出了检验统计量 F_0 的参考分布 ($F_{3,16}$). 显然, 本例中 P 值非常小. 因为 $F_{0.01,3,16} = 5.29$ 且 $F_0 > 5.29$, 所以可以得出结论: P 值的一个上界为 0.01, 即 $P < 0.01$(精确 P 值为 $P = 2.88 \times 10^{-9}$).

表 3.4 晶片蚀刻实验的方差分析表

方差来源	平方和	自由度	均方	F_0	P 值
RF 功率	66 870.55	3	22 290.18	$F_0 = 66.80$	< 0.01
误差	5 339.20	16	333.70		
总和	72 209.75	19			

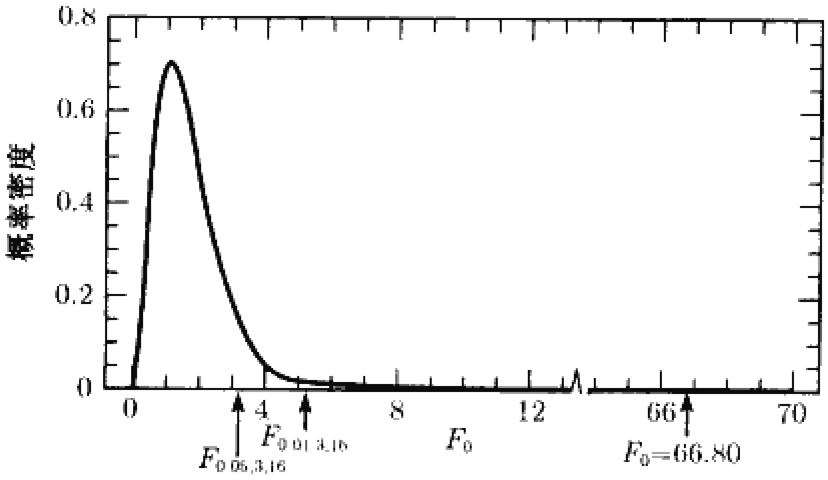


图 3.3 例 3.1 中检验统计量 F_0 的参考分布 ($F_{3,16}$)

1. 关于手工计算的更多内容

读者可能注意到我们是用均值定义平方和的. 也就是说, 由 (3.6) 式,

$$SS_{处理} = n\sum_{i=1}^{a} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

但是我们得到了利用总和的计算公式. 例如, 为了计算 $SS_{处理}$, 我们利用 (3.9) 式:

$$SS_{处理} = \frac{1}{n}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N}$$

理由是这样简单, 并且, $y_{i.}$ 与和 $y_{..}$ 不像均值 $\bar{y}_{i.}$ 和 $\bar{y}_{..}$ 那样容易受舍入误差 (rounding error) 的影响.

通常, 我们不必太关心计算问题, 因为有很多计算机程序可用来进行计算. 这些计算机程序亦有助于很多与实验设计有关的其他分析 (如残差分析和模型适合性检验). 在很多情况下, 这些程序也有助于实验者进行设计.

当有必要手算时, 对观测值进行规范化有时是有用的. 这一点将在下面的例子中加以说明.

例 3.2 对观测值进行规范化

在方差分析的计算中, 对观测值进行**规范化**常会使得计算更简单或更精确. 例如, 考虑例 3.1 中的晶片蚀刻数据. 设对每个观测值都减去 600. 已规范的数据见表 3.5. 容易验证

$$SS_T = (-25)^2 + (-58)^2 + \cdots + (110)^2 - (355)^2/20 = 72\ 209.75$$

$$SS_{处理} = \frac{1}{5}[(-244)^2 + (-63)^2 + (127)^2 + (535)^2] - (355)^2/20 = 66\ 870.55$$

$$SS_E = 5\ 339.20$$

将这些平方和与从例 3.1 中所得的平方和进行比较可以看出, 从原始数据中减去一个常数并不改变平方和.

表 3.5 例 3.2 已规范化的蚀刻率数据

RF 功率 (W)	观测值					总和
	1	2	3	4	5	$y_{i.}$
160	−25	−58	−70	−61	−30	−244
180	−35	−7	−10	−21	10	−63
200	0	51	10	37	29	127
220	125	100	115	85	110	535

今假定对例 3.1 中的每个观测值乘以 2. 容易验证, 变换后的数据的平方和是 SS_T=288 839.00, $SS_{处理}$=267 482.20, 以及 SS_E=21 356.80, 这些平方和明显不同于从例 3.1 中得到的. 不过, 如果将它们同除以 4(即, 2^2), 结果则是相同的. 例如, 处理平方和 267 482.20/4=66 870.55. 还有, 对已规范的数据, F 的比值是 F=(267 482.20)/(21 356.80)=66.80, 也与原始数据的 F 的比值相等. 这样一来, 方差分析是等价的.

2. 随机化检验法与方差分析

在方差分析的 F 检验法研究中, 我们曾经用到了这一假定, 即随机误差 ε_{ij} 是正态独立分布的随机变量. F 检验法也可被证明是**随机化检验法**的一种近似方法. 为了说明这一点, 假定两种处理的每一种处理有 5 个观测值并且我们想检验它们的处理均值是否相等. 数据如下:

处理 1	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}
处理 2	y_{21}	y_{22}	y_{23}	y_{24}	y_{25}

可以用方差分析的 F 检验法来检验 $H_0: \mu_1 = \mu_2$, 也可以用另一种稍微不同的方法来代替. 假设我们考虑把上例中 10 个数据分配给两种处理的所有可能方式. 10 个观测值有 10!/5!5!=252 种可能的配置. 如果处理均值不存在差别, 所有这 252 种配置很可能是相当的. 对于这 252 种配置的每一个配置, 用 (3.7) 式计算 F 统计量的值. 这些 F 值的分布叫做**随机化分布**(randomization distribution), 而一个大的 F 值表示数据与假设 $H_0: \mu_1 = \mu_2$ 不相容. 例如, 如果实际观测到的 F 值仅仅被随机化分布的 5 个值所超过, 则这一点就对应于以显著性水平 $\alpha = 5/252 = 0.019\ 8$(或 1.98%) 拒绝 $H_0: \mu_1 = \mu_2$. 注意, 在这种方法中并不需要有正态性假定.

这种方法的难点是, 即使是对于相对小的问题, 我们也很难计算出确切的随机化分布. 然而, 大量的研究表明: 确切的随机化分布可以用正规理论 F 分布来很好地近似. 这样, 即使没有正

态性假定, F 检验法也可看作为随机化检验法的一种近似方法. 要进一步阅读方差分析中的随机化检验法, 可参阅 Box, Hunter, and Hunter(1978).

3.3.3 模型参数的估计

我们现在给出单因子模型

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

的参数估计和处理均值的置信区间. 随后将证明总均值和处理效应的合理估计可由下式给出:

$$\hat{\mu} = \overline{y}_{..}$$
$$\hat{\tau}_i = \overline{y}_{i.} - \overline{y}_{..}, \quad i = 1, 2, \cdots, a \tag{3.11}$$

这些估计量有相当直观的意义, 注意总均值可以用观测值的总平均来估计, 而任一处理效应只不过是处理均值和总均值之差.

容易确定第 i 个处理均值的**置信区间**(confidence interval). 第 i 个处理的均值是

$$\mu_i = \mu + \tau_i$$

μ_i 的一个点估计是 $\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \overline{y}_{i.}$. 现在, 如果假定误差是正态分布的, 则每一个 $\overline{y}_{i.}$ 是 $\text{NID}(\mu_i, \sigma^2/n)$. 于是, 如果 σ^2 已知, 则我们可用正态分布来确定置信区间. 用 MS_E 作为 σ^2 的估计, 就可以根据 t 分布来建立置信区间. 因此, 第 i 个处理均值 μ_i 的 $100(1-\alpha)\%$置信区间是

$$\overline{y}_{i.} - t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}} \leqslant \mu_i \leqslant \overline{y}_{i.} + t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}} \tag{3.12}$$

任意两个处理均值之差 (即 $\mu_i - \mu_j$) 的一个 $100(1-\alpha)\%$置信区间是

$$\overline{y}_{i.} - \overline{y}_{j.} - t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}} \leqslant \mu_i - \mu_j \leqslant \overline{y}_{i.} - \overline{y}_{j.} + t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}} \tag{3.13}$$

例 3.3 用例 3.1 的数据, 得到总均值的估计量和处理效应的估计量为 $\hat{\mu} = 12\ 355/20 = 617.75$ 以及

$$\hat{\tau}_1 = \overline{y}_{1.} - \overline{y}_{..} = 551.20 - 617.75 = -66.55$$
$$\hat{\tau}_2 = \overline{y}_{2.} - \overline{y}_{..} = 587.40 - 617.75 = -30.35$$
$$\hat{\tau}_3 = \overline{y}_{3.} - \overline{y}_{..} = 625.40 - 617.75 = 7.65$$
$$\hat{\tau}_4 = \overline{y}_{4.} - \overline{y}_{..} = 707.00 - 617.75 = 89.25$$

处理 4(RF 功率为 220W) 的均值的一个 95%置信区间由 (3.12) 式算得是

$$707.00 - 2.120\sqrt{\frac{333.70}{5}} \leqslant \mu_4 \leqslant 707.00 + 2.120\sqrt{\frac{333.70}{5}}$$
$$707.00 - 17.32 \leqslant \mu_4 \leqslant 707.00 + 17.32$$

于是, 所求得的置信区间是 $689.68 \leqslant \mu_4 \leqslant 724.32$.

联合置信区间

(3.12) 式和 (3.13) 式给出的置信区间表达式是**一次一个**(one-at-a-time) 的置信区间. 也就是, 置信水平 $1-\alpha$ 只应用于一个特定的估计. 但是, 在许多问题中, 实验者希望计算多个置信区间, 每个均值或均值差都对应一个置信区间. 如果有这样 r 个感兴趣的 $100(1-\alpha)\%$的置信区间, r 个置信区间**同时**正确的概率至少为 $1 - r\alpha$. 概率 $r\alpha$ 通常被称为**实验方法误差率**或总的

置信系数. 与其众多置信区间提供相对很少的信息, 宁可区间数 r 不要太大. 例如有 $r=5$ 个区间和 $\alpha=0.05$(典型的选择), 5 个置信区间的联合置信水平至少为 0.75, 如果 $r=10$ 个区间和 $\alpha=0.05$, 联合置信水平至少为 0.50.

确保联合置信水平不太小的方法是在一次一个的置信区间 (3.12) 式和 (3.13) 式中用 $\alpha/(2r)$ 替代 $\alpha/2$. 这方法叫**Bonferroni 方法**, 它允许实验者构造一列 r 个处理均值或处理均值差异的联合置信区间, 使得总体置信水平至少为 $100(1-\alpha)\%$. 当 r 不太大时, 合理地缩短置信区间长度是一个非常好的方法. 详见第 3 章的**补充材料**.

3.3.4 不平衡数据

在有些单因子实验中, 每种处理所取的观测值个数可以是各不相同的. 我们称这样的设计是**不平衡的**. 稍许修改一下平方和公式, 上面所述的方差分析法仍可使用. 设在处理 $i(i=1,2,\cdots,a)$ 中取 n_i 个观测值, 令 $N=\sum_{i=1}^{a} n_i$. 关于 SS_T 和 $SS_{处理}$ 的计算公式变为

$$SS_T=\sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2-\frac{y_{..}^2}{N} \tag{3.14}$$

与

$$SS_{处理}=\sum_{i=1}^{a}\frac{y_{i.}^2}{n_i}-\frac{y_{..}^2}{N} \tag{3.15}$$

在方差分析中无需其他变更.

选取平衡设计有两个优点. 首先, 相对来说, 当样本量相等时, 检验统计量对 a 个处理等方差的假定的微小偏离是不敏感的. 但对不相等样本量的情况则并非如此. 其次, 当样本量相等时, 检验的功效最大.

3.4 模型适合性检验

由方差分析恒等式 (3.6) 给出的观测值变异性的分解式是一个纯代数的关系式. 而利用这一分解法来正式地检验处理均值之间没有差异是需要满足一定的假设条件的. 具体说, 这些假设条件是观测值能够用模型

$$y_{ij}=\mu+\tau_i+\varepsilon_{ij}$$

恰当地描述, 并且, 这些误差服从正态独立分布, 其均值为零, 方差为未知常数 σ^2. 如果这些假定条件有效, 则方差分析方法是处理均值没有差异这一假设的一种较为理想的检验法.

然而, 在实际当中, 这些假定条件通常不完全成立. 因此, 在这些假定条件的有效性尚未核实之前就依靠方差分析, 通常是不明智的. 是否违背了这些基本假定和模型的适合性, 可以很容易地利用**残差**检测来进行分析. 处理 i 的观测值 j 的残差定义为

$$e_{ij}=y_{ij}-\hat{y}_{ij} \tag{3.16}$$

其中 $\hat{y}_{ij}$ 是对应于 y_{ij} 的一个估计, 由下式得出:

$$\hat{y}_{ij}=\hat{\mu}+\hat{\tau}_i=\overline{y}_{..}+(\overline{y}_{i.}-\overline{y}_{..})=\overline{y}_{i.} \tag{3.17}$$

(3.17) 式直观地给出了引人瞩目的结果, 即第 i 个处理任一观测值的估计恰好是对应的处理平均值.

残差检验是任一方差分析不可缺少的部分. 如果模型是适合的, 则残差是**无定形的**. 也就是说, 它们没有明显的模式. 通过研究残差, 可以发现很多模型不适合和基本假定不符合的例子. 在这部分, 我们将说明如何通过采用残差的图形分析来很容易地进行模型诊断检测, 以及如何处置几种常见的反常现象.

3.4.1 正态性假设

检验正态性假设可以利用残差直方图. 如果满足关于误差的 $\text{NID}(0, \sigma^2)$ 假定, 则此直方图就应该类似于取自中心在零点处的正态分布的样本. 可惜, 对于小样本, 经常会出现明显的波动, 所以图像上中度偏离正态性的出现并不一定意味着严重违背正态性假定. 而图像上对于正态性的严重偏离则有严重违背正态性假定的可能, 并需要进行进一步的分析.

另一种极其有用的方法是构造一个残差的**正态概率图**. 回顾第 2 章的 t 检验, 我们用原始数据画成的正态概率图来检验正态性假设. 在方差分析中, 对残差这样做通常更有效 (且更直接). 如果潜在的误差分布是正态的, 则图像呈直线状. 在想象这一直线时, 与极端值点相比, 我们应该更看重直线附近的值点.

表 3.6 列出了例 3.1 中的蚀刻速率数据的原始数据和残差. 图 3.4 是残差的正态概率图, 审视这个图形, 总的印象是误差分布是近似正态的. 正态概率图在左边稍下弯曲, 在右边稍有上翘, 这意味着误差分布的尾部比起正态分布的尾部要更细一些; 也就是说, 最大的残差不完全如所期望的那样大 (在绝对值的意义下). 但是, 此图总体看起来却是正态的.

表 3.6 例 3.1 的蚀刻速率数据和残差 [a]

功率(W)	观测(j) 1	2	3	4	5	$\hat{y}_{ij}=\bar{y}_{i.}$
160	23.8 575 (13)	−9.2 542 (4)	−21.2 530 (14)	−12.2 539 (8)	18.8 570 (5)	551.2
180	−22.4 565 (17)	5.6 593 (6)	2.6 590 (18)	−8.4 579 (16)	22.6 610 (9)	587.4
200	−25.4 600 (20)	25.6 651 (1)	−15.4 610 (19)	11.6 637 (7)	3.6 629 (10)	625.4
220	18.0 725 (2)	−7.0 700 (15)	8.0 715 (11)	−22.0 685 (12)	3.0 710 (3)	707.0

a 每单元的盒内是残差. 圆括号内的数表示收集数据的序号.

一般地, 在固定效应方差分析中适度地偏离正态性我们并不大在意 (回顾在 3.3.2 节中对随机化试验的讨论). 比起偏态分布来说, 我们更为关心的是比正态分布的尾部明显的厚或薄的误差分布. 因为 F 检验法只受轻微的影响, 所以我们说, 方差分析 (及其有关的方法, 例如多重比较法) 对正态性假定是**稳健**的. 偏离正态性通常会引起真正的显著性水平和功率与名义上的数值稍有差异, 一般来说, 功率会偏低. 我们将在第 12 章中讨论的随机效应模型较多地受到非正

态性的影响.

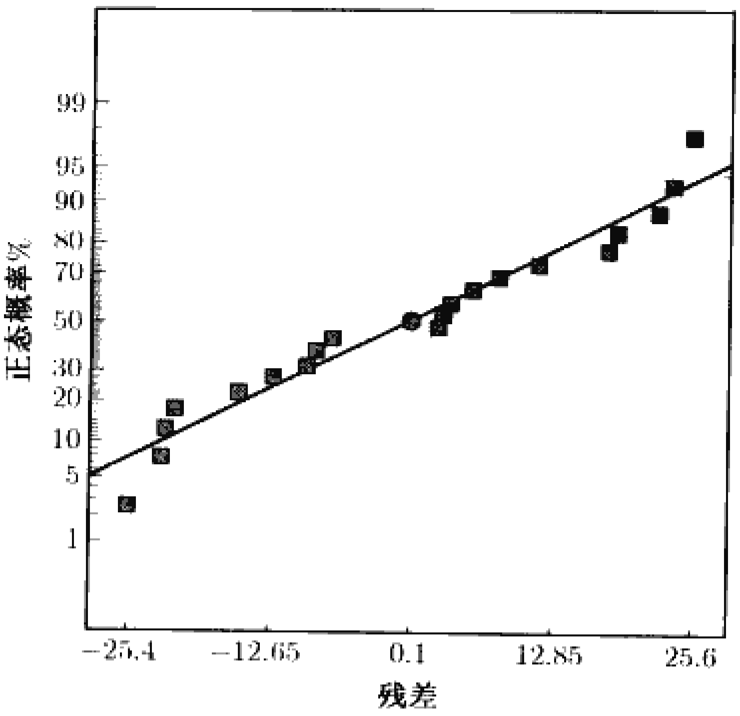


图 3.4　例 3.1 的残差的正态概率图

正态概率图上的一种十分普遍的毛病, 是有一个残差远大于其他的残差. 这样的一个残差通常叫做**异常值**(outlier). 一个或多个异常值的出现会严重干扰方差分析, 所以, 要小心探究出现的异常值. 引起异常值的原因经常是由于计算发生错误、规范数据或复制数据造成的误差所致. 否则, 则必须仔细研究该实验周围的实验环境. 如果异常的响应是一个特别希望得到的值 (高强度, 低价格, 等等), 则异常值比其他数值更为有用. 不要轻易拒绝或放弃一个异常的观测值, 除非有合理的非统计性的根据. 就是在最不利的情况下, 也要做两次分析, 一次包括异常值, 另一次剔除异常值.

有几种正式的统计方法可用来检测异常值 [例如, 参阅 Barnett and Lewis (1994), John and Prescott(1975), 以及 Stefansky(1972)]. 粗略检测异常值可以由验算**标准化残差**

$$d_{ij} = \frac{e_{ij}}{\sqrt{MS_E}} \tag{3.18}$$

来做到. 如果误差 ε_{ij} 服从 $N(0,\sigma^2)$, 则标准化残差应是渐近正态分布的, 其均值为零, 方差为 1. 于是, 标准化残差约 68%落在 ±1 之内, 约 95%落在 ±2 之内, 几乎全部落在 ±3 之内. 比零点要大于 3 或 4 倍标准差的残差是一个可能的异常值.

对于例 3.1 的蚀刻速率数据, 正态概率图没有显出异常值. 还有, 最大的标准化残差是

$$d_1 = \frac{e_1}{\sqrt{MS_E}} = \frac{25.6}{\sqrt{333.70}} = \frac{25.6}{18.27} = 1.40$$

但它并不重要.

3.4.2　依时间序列的残差图

依照收集数据的时间顺序画出残差图有助于检测残差之间的**相关性**. 具有正残差和负残差的趋势表明了正相关性. 而这说明不符合误差的**独立性假定**(independence assumption). 这是

一个潜在的严重问题且难于校正的问题, 所以, 要在收集数据时尽可能地防止这一问题的发生. 而实验的适当随机化是获得独立性的一个重要步骤.

有时, 实验者 (或主体) 的技巧可能会在实验进程中改变, 或者, 所研究的过程可能 "漂移" 或变得更加不稳定. 这经常会导致误差方差随时间而改变. 此条件常使残差关于时间的图形在一边比另一边更为伸展. 非常数方差是一个潜在的严重问题. 我们将在 3.4.3 节和 3.4.4 节中更多地讨论这个主题.

表 3.6 列出了例 3.1 中蚀刻速率数据的残差和收集数据的时间序列. 残差与试验次序或时间的关系图如图 3.5 所示. 没有理由去怀疑存在违反独立性或常数方差的假定.

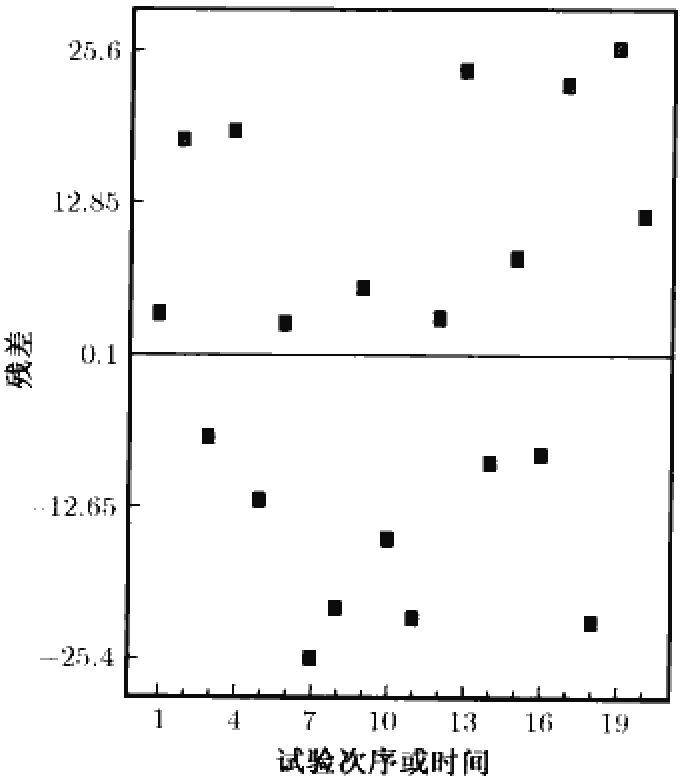


图 3.5 试验次序或时间的残差图

3.4.3 残差与拟合值的关系图

如果模型是正确的并且满足假定的条件, 则残差应该是无定形的. 特别地, 它们应该与任一其他变量没有任何关系, 自然也与用来预测响应的变量无关. 一种简单的检测法是画出残差与拟合值 $\hat{y}_{ij}$ 的关系图. [对单因子实验模型来说, $\hat{y}_{ij} = \bar{y}_{i.}$(即第 i 种处理的平均值).] 该图应不显现出任何明显的模式. 图 3.6 是例 3.1 中蚀刻速率数据的残差与拟合值的关系图. 没有出现异常的结构.

从这种图形上偶尔会检测出**非常数的方差**. 有时, 观测值的方差会随着观测值数量的增加而增加. 如果实验中的误差或背景噪音所占观测值大小的比例是一常数的话, 这种情况就会发生. (这通常出现在测量工具的误差是读数尺寸的百分比的时候.) 万一出现这种情况, 那么残差会随着 y_{ij} 的增大而增大, 而残差与 $\hat{y}_{ij}$ 的关系图看起来就像一个向外开口的漏斗或喇叭筒. 当数据服从非正态的偏态分布时也会出现非常数方差, 因为在偏态分布中, 方差大体上是均值的函数.

如果违反了方差的齐性假定, 则 F 检验法在平衡 (所有处理的样本量相等) 固定效应模型中只会受到轻微的影响. 但是, 在不平衡设计或一个方差远大于其他方差的情形中, 问题就比较严重了. 特别地, 如果因子水平有很大的方差, 同时有较小的样本量, 则实际的第一类错误率远

大于期望值 (或置信水平比预期的低). 相反, 如果因子水平有很大的方差, 同时有较大的样本量, 则显著水平比预期的要小 (置信水平比预期的高). 这是尽可能选择**等样本量** (equal sample sizes) 的一个充分理由. 对于随机效应模型, 不相等的误差方差可能会显著地干扰关于方差分量的推断, 即使用平衡设计也是如此.

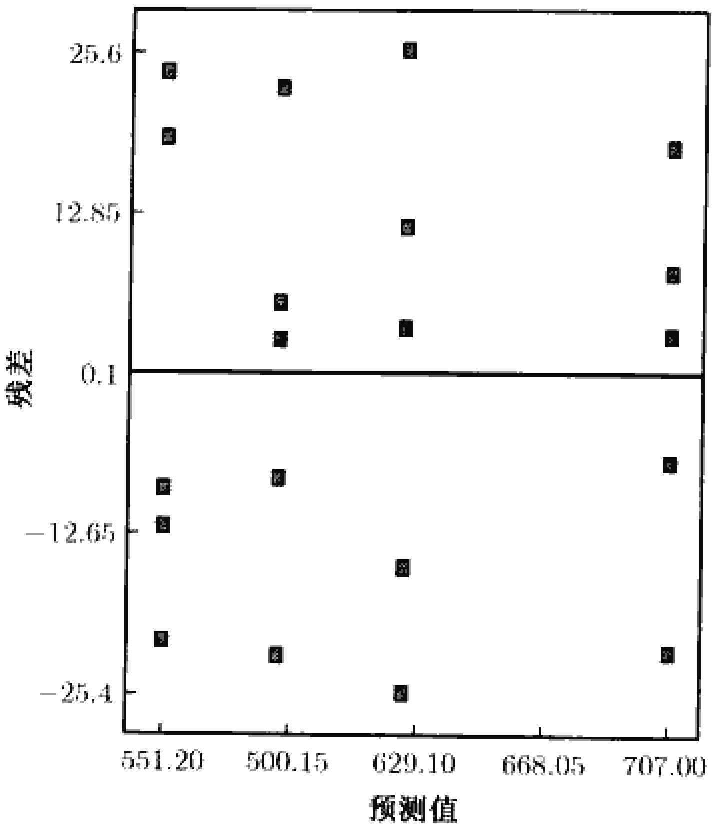


图 3.6　残差与拟合值的关系图

当非常数方差是因为上面的理由而出现的时候, 通常的处理方案是应用**方差稳定化变换** (variance-stabilizing transformation), 然后对变换后的数据进行方差分析. 注意在这种处理方案中, 我们是对经过变换后的总体应用方差分析的结论.

关于如何选取恰当的变换已有相当多的研究. 如果实验者知道观测值的理论分布, 他们就可以利用这一信息来选取变换. 例如, 如果观测值服从泊松分布, 则可以用**平方根变换** $y_{ij}^* = \sqrt{y_{ij}}$ 或 $y_{ij}^* = \sqrt{1+y_{ij}}$. 如果数据服从对数正态分布, 则可用**对数变换** $y_{ij}^* = \ln y_{ij}$. 对用分数表示的二项分布数据, 则用**反正弦变换** $y_{ij}^* = \arcsin\sqrt{y_{ij}}$. 在没有明显的变换时, 实验者通常凭经验来寻找使方差不变的变换, 而不管均值的数值如何. 在第 5 章所讨论的析因实验中, 另一种处理方案是选取使交互作用的均方取最小值的变换, 使实验所得的结果较易于解释. 在第 15 章中, 我们将较详细地讨论选取变换的解析方法. 对方差的不等性所作的变换也会影响到误差分布的形式. 在大多数情况下, 变换带来的误差分布更接近于正态分布. 关于变换的更详细的讨论, 请参阅 Bartlett(1947), Box and Cox(1964), Dolby(1963), Draper and Hunter(1969).

1. *方差相等性的统计检验法*

残差图可经常用来诊断方差的不相等性, 不过也可运用统计检验法 (有好几种). 这些检验法可以从形式上看作是检验假设

$$H_0:\ \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2, \qquad H_1:\ \text{至少对一个 } \sigma_i^2 \text{ 不真}$$

一个广为应用的方法是 **Bartlett 检验法**. 此方法涉及计算一个统计量, 当随机样本来自独立正态总体时, 其抽样分布渐近于自由度为 $a-1$ 的卡方分布. 检验统计量是

$$\chi_0^2 = 2.302\,6\ \frac{q}{c} \tag{3.19}$$

其中

$$q = (N-a)\lg S_p^2 - \sum_{i=1}^{a}(n_i-1)\lg S_i^2$$

$$c = 1 + \frac{1}{3(a-1)}\left(\sum_{i=1}^{a}(n_i-1)^{-1} - (N-a)^{-1}\right)$$

$$S_p^2 = \frac{\sum_{i=1}^{a}(n_i-1)S_i^2}{N-a}$$

而 S_i^2 是第 i 个总体的样本方差.

当所有的 S_i^2 相等时, 量 q 等于零, 当样本方差 S_i^2 有较大的差异时, 量 q 较大. 因此, 当 χ_0^2 的值太大时, 我们应当拒绝 H_0. 也就是说, 仅当

$$\chi_0^2 > \chi_{\alpha,a-1}^2$$

时, 拒绝 H_0, 其中 $\chi_{\alpha,a-1}^2$ 是自由度为 $a-1$ 的卡方分布的上 α 百分位数. 决策时也可以利用 P 值.

Bartlett 检验法对正态性假定非常敏感, 当怀疑正态性假定时, 最好不要使用 Bartlett 检验法.

例 3.4 因为正态性假定不成问题, 我们将 Bartlett 检验法应用于例 3.1 中的蚀刻速率实验数据. 首先计算每种处理的样本方差, 求得 $S_1^2 = 400.7$, $S_2^2 = 280.3$, $S_3^2 = 421.3$, $S_4^2 = 232.5$. 则

$$S_p^2 = \frac{4(400.7)+4(280.3)+4(421.3)+4(232.5)}{16} = 333.7$$

$$q = 16\lg(333.7) - 4[\lg 400.7 + \lg 280.3 + \lg 431.3 + \lg 232.5] = 0.17$$

$$c = 1 + \frac{1}{3(3)}\left(\frac{4}{4} - \frac{1}{16}\right) = 1.10$$

检验统计量是

$$\chi_0^2 = 2.302\,6\frac{(0.17)}{(1.10)} = 0.36$$

因为 $\chi_{0.05,3}^2 = 7.81$, 所以不能拒绝零假设, 且结论是, 所有 4 个方差是相同的. 分析残差与拟合值的关系图也能得到同样的结论.

因为 Bartlett 检验法对正态性假定非常敏感, 所以提出另一种方法也许是合适的. Anderson and McLean(1974) 很好地讨论了关于方差相等性的统计检验. **修正后的 Levene 检验**[见 Levene(1960) 和 Conover, Johoson, and Johoson(1981)] 是一种非常好的方法, 它对正态性的偏离是稳健的. 为了检验所有处理的等方差假设, 修正后的 Levene 检验利用了每个处理的观测值 y_{ij} 与该处理的中位数 (记 $\tilde{y}_{i.}$) 的偏差的绝对值, 用

$$d_{ij} = |y_{ij} - \tilde{y}_{i.}| \qquad \begin{cases} i = 1,2,\cdots,a \\ j = 1,2,\cdots,n_i \end{cases}$$

来表示这些偏差. 修正后的 Levene 检验评估了所有处理的平均偏差是相等的还是不等的. 如果这些平均偏差是相等的, 则所有处理中的观测值的方差将是相同的. Levene 检验的检验统计量只是简单地把通常用于检验均值相等的 ANOVA F 统计量用到了绝对偏差上.

例 3.5 一位土木工程师试图确定: 4 种不同的洪水流量频率的估计方法在应用于同一流域时, 是否能产生相同的峰值排水量估计. 每种方法在此流域中使用 6 次, 所得的排水量数据 (单位为每秒立方英尺) 如表 3.7 的上半部所示. 数据的方差分析概括在表 3.8 中, 它表明 4 种方法给出的洪峰排水量的均值不同. 残差与拟合值的关系图如图 3.7 所示, 令人不爽的是, 开口向外的漏斗型表明所得数据不满足常值方差的假定.

表 3.7 峰值排水量数据

估计方法	观测值						$\bar{y}_{i.}$	$\tilde{y}_i$	S_i	修正后的 Levene 检验的偏差 d_{ij}					
1	0.34	0.12	1.23	0.70	1.75	0.12	0.71	0.520	0.66	0.18	0.40	0.71	0.18	1.23	0.40
2	0.91	2.94	2.14	2.36	2.86	4.55	2.63	2.610	1.092	1.70	0.33	0.47	0.25	0.25	1.94
3	6.31	8.37	9.75	6.09	9.82	7.24	7.93	7.805	1.66	1.495	0.565	1.945	1.715	2.015	0.565
4	17.15	11.82	10.95	17.20	14.35	16.82	14.72	15.59	2.77	1.56	3.77	4.64	1.61	1.24	1.23

表 3.8 峰值排水量数据的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
方法	708.347 1	3	236.115 7	76.07	< 0.001
误差	62.081 1	20	3.104 1		
总和	770.428 2	23			

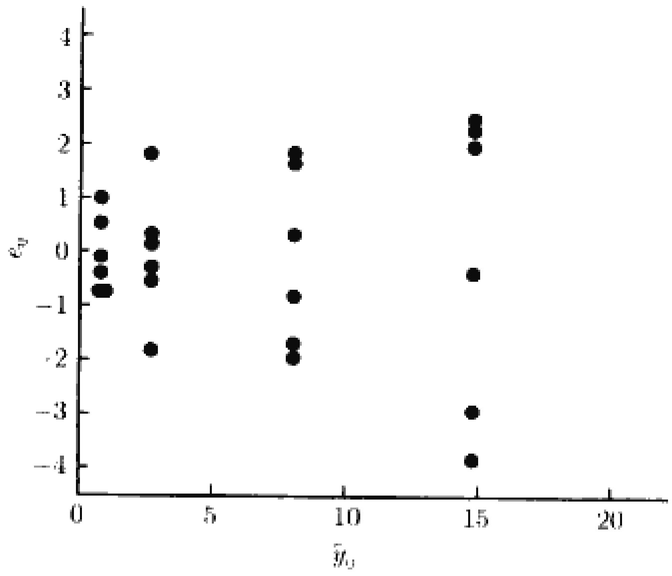


图 3.7 例 3.5 中残差与拟合值的关系图

我们用修正后的 Levene 方法检验峰值排水量数据. 表 3.7 的左半部含有处理的中位数 $\tilde{y}_i$, 而右半部含有对中位数的偏离 d_{ij}. Levene 检验进行 d_{ij} 的方差标准分析. 得到的 F 检验统计量 F_0 =4.55, 其 P 值是 P =0.013 7. 因此, Levene 检验拒绝了方差相等的零假设. 我们的结论与图 3.7 所显示的信息一致. 峰值排水量数据是数据变换的一个好的例子.

2. 变换的经验选择

正如前面我们观测到的, 如果实验者了解观测的方差和均值之间的关系, 则他们可以利用这一信息来选择变换的形式. 现在, 我们进一步说明这一点, 并指出怎样做到由数据来选择所需要的变换形式.

设 $E(y) = \mu$ 是 y 的均值, 假定 y 的标准差与 y 的均值的幂成正比, 即

$$\sigma_y \propto \mu^{\alpha}$$

我们想寻找 y 的一种变换使它能够产生一个常数方差. 假定这个变换是原始数据的幂, 比方说

$$y^* = y^\lambda \tag{3.20}$$

则它可以表示成

$$\sigma_{y^*} \propto \mu^{\lambda+\alpha-1} \tag{3.21}$$

显然, 如果设 $\lambda = 1-\alpha$, 则变换后的数据 y^* 的方差是常量.

前面讨论的几种常用的变换概括在表 3.9 中. 注意, $\lambda = 0$ 意味着是对数变换. 这些变换按照强度的递增顺序排列. 所谓变换的强度, 是指它引起的曲率. 应用于一个窄区间上的数据的弱变换在分析上的效应不大, 而应用于大区间上的强变换则会有戏剧性的结果. 除非比值 $y_{\max}/y_{\min}$ 大于 2 或 3, 否则变换效果通常并不明显.

表 3.9 方差稳定化变换

σ_y 和 μ 间的关系	α	$\lambda = 1-\alpha$	变 换	备 注
$\sigma_y \propto$ 常数	0	1	没有变换	
$\sigma_y \propto \mu^{1/2}$	1/2	1/2	平方根	泊松 (记数) 数据
$\sigma_y \propto \mu$	1	0	对数	
$\sigma_y \propto \mu^{3/2}$	3/2	−1/2	平方根的倒数	
$\sigma_y \propto \mu^2$	2	−1	倒数	

在许多有重复的实验设计情形中, 我们可以凭经验根据数据估出 α. 因为在第 i 个处理组合中 $\sigma_{y_i} \propto \mu_i^\alpha = \theta\mu_i^\alpha$, 这里 θ 是比例常数, 取对数, 有

$$\lg \sigma_{y_i} = \lg\theta + \alpha \lg \mu_i \tag{3.22}$$

因此, $\lg\sigma_{y_i}$ 与 $\lg\mu_i$ 的关系图是斜率为 α 的一条直线. 因为不知道 σ_{y_i} 和 μ_i, 我们也许可以把它们合理的估计代入 (3.22) 式中, 用拟合直线的斜率来作为 α 的估计值. 典型地, 我们可以用第 i 个处理 (或者, 更一般地, 第 i 个处理组合或第 i 组实验条件) 的标准差 S_i 和均值 $\overline{y}_{i.}$ 来估计 σ_{y_i} 和 μ_i.

为了研究对例 3.5 中的峰值排水量数据采用方差稳定化变换的可能性, 我们在图 3.8 中画出 $\lg S_i$ 与 $\lg\overline{y}_{i.}$ 的关系图. 过这 4 点的直线斜率接近 1/2, 从表 3.9 中可知平方根变换是合适的. 变换后数据 $y^* = \sqrt{y}$ 的方差分析见表 3.10, 残差与预期响应的关系图见图 3.9. 与图 3.7 相比, 此残差图大为改善, 所以我们得出结论, 平方根变换是有帮助的. 注意, 表 3.10 中误差的自由度与总和的自由度都减少了 1, 这是因为在估计变换参数 α 时已利用了这些数据.

实际上, 许多实验者通过简单地尝试几个变换, 观测每种变换在残差与预测值的关系图上的效果, 就可以选择变换的形式. 从而选出能得到最令人满意的残差图的变换.

3.4.4 残差与其他变量的关系图

如果数据收集时还跟其他可能影响响应的变量有关, 则应画出残差与那些变量的关系图. 例如, 在例 3.1 的蚀刻速率实验中, 蚀刻速率可能受气压的显著影响, 所以, 应该画出残差与气压的关系图. 如果用不同的蚀刻机来收集数据, 则应画出残差与那些机器的关系图. 只要在这类残

差中呈现一定的模式, 就意味着相应变量影响响应. 对于这种变量或者在将来的实验中更谨慎地控制, 或者将它也包括在数据分析中.

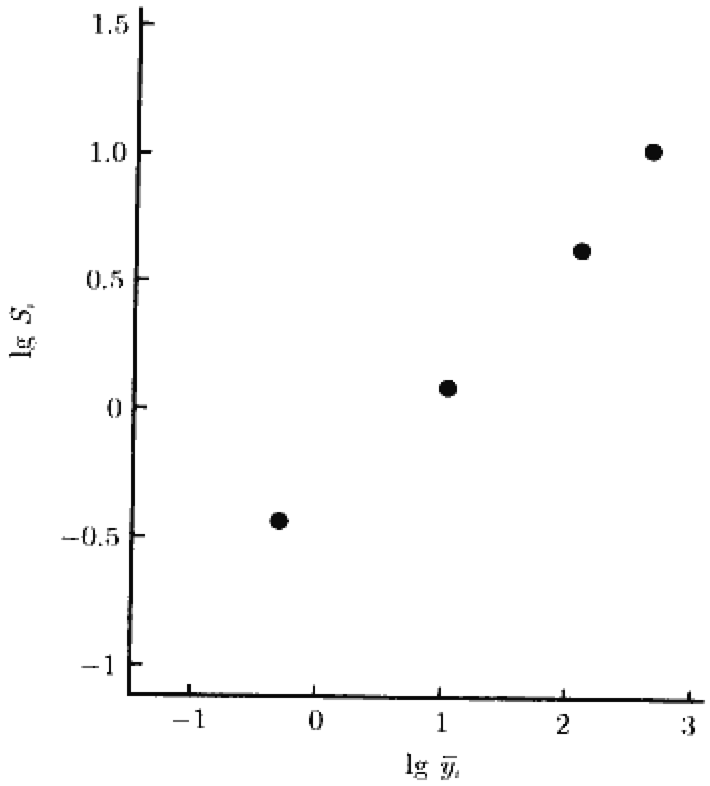


图 3.8 例 3.5 中峰值排水量数据 $\lg S_i$ 与 $\lg \bar{y}_{i.}$ 的关系图

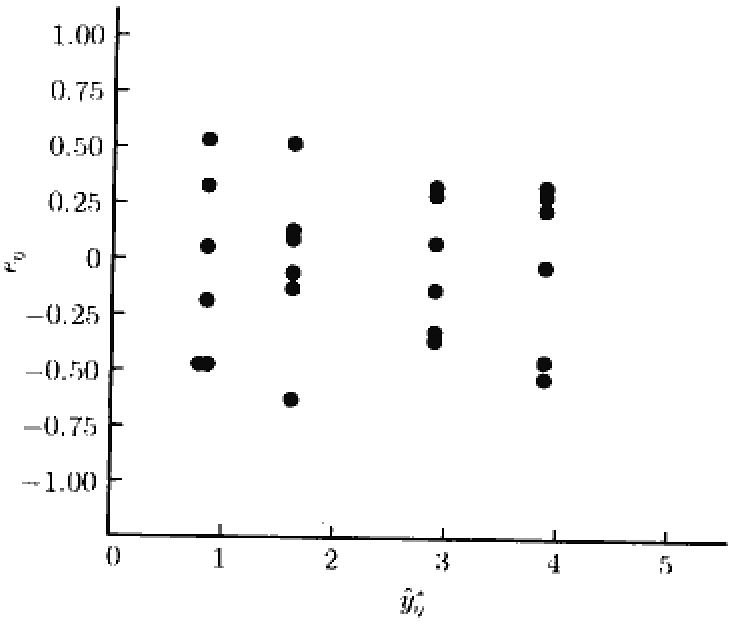


图 3.9 例 3.5 中峰值排水量变换数据的残差与 $\hat{y}^*_{ij}$ 的关系图

表 3.10 变换后峰值排水量数据 $y^* = \sqrt{y}$ 的方差分析

方差来源	平方和	自由度	均　方	F_0	P 值
方法	32.684 2	3	10.894 7	76.99	< 0.001
误差	2.688 4	19	0.141 5		
总和	35.372 6	22			

3.5 结果的实际解释

在完成实验、统计分析以及研究假设后, 实验者将准备针对正研究的问题作出结论. 通常这是相对容易的, 迄今为止, 我们所考虑的简单实验中, 也许通过非规范的方法就可以得出结论, 诸如图 3.1 和图 3.2 中的盒图和散点图的图示检查. 但是, 在某些情形下, 需要应用更多的规范技术. 我们将在本节中部分介绍这些技术.

3.5.1 回归模型

实验中的因子可以是**定量的** (quantitative), 也可以是**定性的** (qualitative). 定量因子的水平与数值尺度的点相对应, 例如温度、压强或时间. 另一方面, 定性因子的水平却不能依量的大小顺序来排列. 操作者、原料的批次以及班次都是典型的定性因子, 因为没有理由将它们依照任一特定的数值顺序来排序.

就最初的实验设计与分析而论, 两种类型的因子是同等对待的. 实验者的兴趣在于确定因

子的水平之间可能存在的差异. 如果因子是定性的, 例如操作者, 则考虑下一次实验中该因子在中间水平处的响应是没有意义的. 但是, 对于定量的因子, 例如时间, 实验者通常感兴趣于取值的整个范围, 特别是下一次实验中处于中间因子水平的响应. 也就是说, 如果在实验中使用了 3 种时间水平, 分别为 1.0 小时、2.0 小时及 3.0 小时, 则我们可能还想预测在 2.5 小时处的响应. 这样一来, 实验者经常想要开发实验中响应变量的内插式. 这个式子就是研究过程的**经验模型** (empirical model).

拟合数据方程的一般方法称为**回归分析** (regression analysis), 这将在第 10 章中详细讨论. 也可以见本章的**补充材料**. 本节用例 3.1 中的蚀刻速率的数据简要说明这种技术.

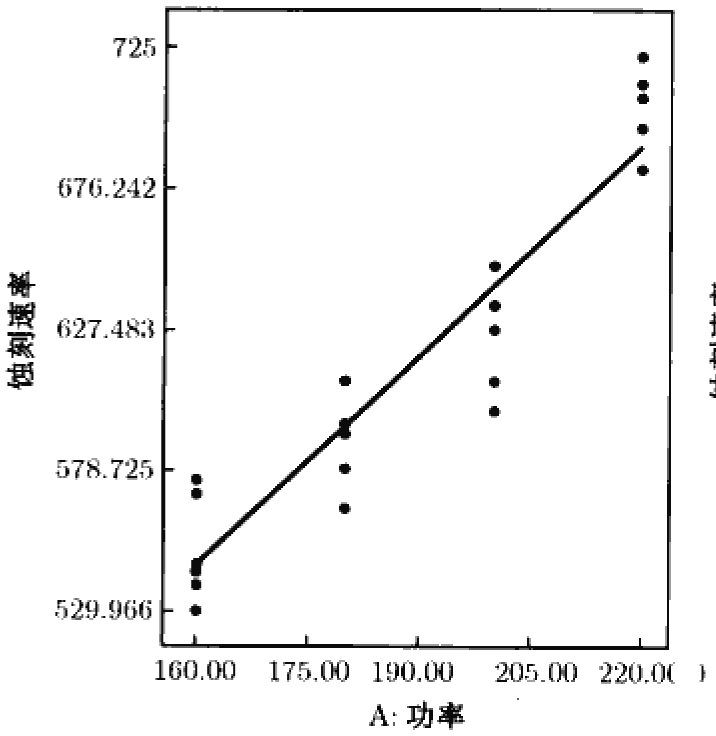
图 3.10 提供了例 3.1 实验中的蚀刻速率 y 关于功率 x 的散点图. 审视这个散点图, 蚀刻速率和功率之间明显存在强烈的相关. 作为第一次近似, 我们用**线性模型**(linear model) 拟合数据, 即

$$y = \beta_0 + \beta_1 x + \varepsilon$$

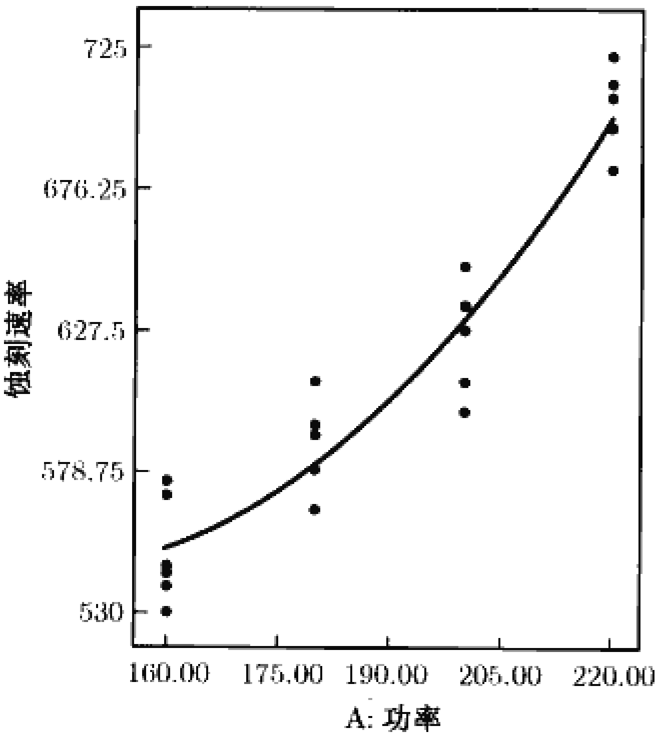
其中 β_0 和 β_1 是要估计的未知参数, ε 是随机误差项. 常用于估计模型中参数的方法是**最小二乘法** (method of least squares). 这包括选择 β 的估计使得误差 (ε) 的平方和最小. 本例中, 最小二乘拟合式是

$$\hat{y} = 137.62 + 2.527x$$

(关于回归方法, 请看第 10 章和本章的补充材料.)



(a) 线性模型



(b) 二次模型

图 3.10 例 3.1 的蚀刻速率数据的散点图和回归模型

此线性模型如图 3.10a 所示. 它在高功率设置下显得不很满意. 或许加进 x 的二次项会得到改善. 所得的**二次模型**拟合的结果是

$$\hat{y} = 1\ 147.77 - 8.255\ 5x + 0.028\ 375x^2$$

此二次拟合式如图 3.10b 所示. 二次模型看上去要优于线性模型, 因为它在高功率设置处拟合得较好.

一般说来, 只要能足够地描述系统或过程, 我们更喜欢用最低阶的多项式去拟合. 本例中, 二次多项式看来比线性模型拟合得好, 所以, 二次模型额外的复杂性是值得的. 选择近似多项式的阶并不是一件容易的事情, 因为比较容易过分拟合, 也就是多加了高阶项. 这不但不能真正改善拟合式, 反而会增加模型的复杂性并且常常会损坏它作为预测值或内插式的可用性.

本例中, 经验模型可以在实验的区域里用来预测功率设置上的蚀刻速率. 在其他情况下, 经验模型可以用于**过程最优化** (process optimization). 也就是, 基于响应的最佳值寻找变量的最佳值. 我们将在本书的后面更加全面地讨论和说明这个问题.

3.5.2 处理均值的比较

假设对固定效应模型引入方差分析时我们拒绝了零假设. 于是, 处理均值之间会存在一定的差异, 但并未确切指出哪些均值不相同. 此时, 进一步在处理均值组之间进行比较和分析可能是有用的. 第 i 个处理均值定义为 $\mu_i = \mu + \tau_i$, 且用 $\bar{y}_{i.}$ 估计 μ_i. 或者用处理总和 $\{y_{i.}\}$, 或者用处理平均值 $\{\bar{y}_{i.}\}$ 来进行处理均值间的比较. 进行这类比较的方法通常叫做**多重比较法** (multiple comparison method). 接下来的几节将讨论在单个处理均值间或处理均值组间的比较方法.

3.5.3 均值的图解比较法

基于方差分析结果, 很容易提出一个图解法进行均值比较. 设所感兴趣的因子有 a 个水平, $\bar{y}_{1.}, \bar{y}_{2.}, \cdots, \bar{y}_{a.}$ 是处理均值. 当 σ 已知时, 任一处理均值的标准差是 $\sigma/\sqrt{n}$. 因此, 当所有因子水平的均值都相等时, 观测得到的样本均值 $\bar{y}_{i.}$ 应当服从均值为 $\bar{y}_{..}$、标准差为 $\sigma/\sqrt{n}$ 的正态分布. 可以想象一个可以沿着数轴移动的正态分布, 轴的下面用点标出 $\bar{y}_{1.}, \bar{y}_{2.}, \cdots, \bar{y}_{a.}$. 如果处理均值都相等, 则分布可以移动到某一位置, 使得 $\bar{y}_{i.}$ 的值来自同一分布. 如果处理均值不全相等, 则由于 $\bar{y}_{i.}$ 的值与产生不同响应均值的因子水平有关, 所以它们不像是来自于同一分布.

逻辑上唯一的缺点是 σ 未知. 不过, 由方差分析可知, 我们可以用 $\sqrt{MS_E}$ 来代替 σ 并用带有尺度因子 $\sqrt{MS_E/n}$ 的 t 分布来代替正态分布. 这样的方法用于例 3.1 中的蚀刻率数据如图 3.11 所示. 注意到图的中央用实线表示 t 分布.

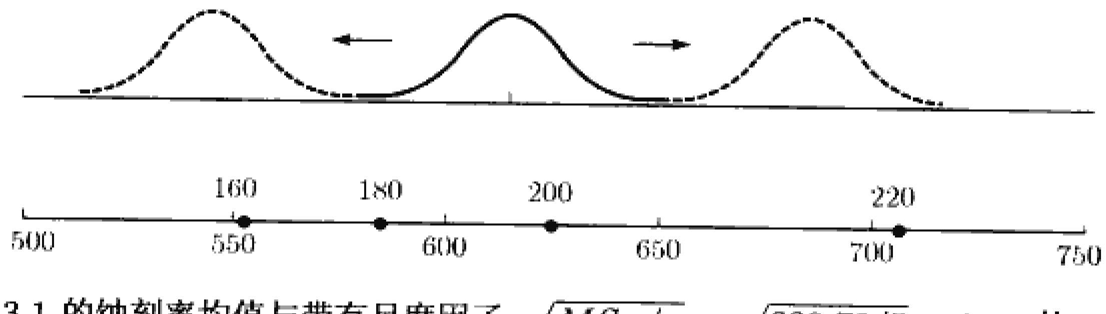


图 3.11 例 3.1 的蚀刻率均值与带有尺度因子 $\sqrt{MS_E/n} = \sqrt{330.70/5} = 8.13$ 的 t 分布的比较图

为了在图 3.11 画出 t 分布的草图, 将横坐标 t 的值乘以尺度因子

$$\sqrt{MS_E/n} = \sqrt{330.70/5} = 8.13$$

并在 t 的纵坐标上描出对应的点即得此草图. 因为 t 分布看起来很像正态分布, 只是在中心附近稍微扁平且尾部较长, 所以, 很容易画出该草图. 若想要更精确一些, 在 Box, Hunter, and Hunter(1978) 的书中有一张 t 的横坐标值和对应的纵坐标的表可供参阅. 分布的坐标原点可任意选择, 通常最好选取 $\bar{y}_{i.}$ 范围中的一个值, 以便于比较. 在图 3.11 中, 原点是 615Å/ min.

现在, 在图 3.11 中, 想象沿着水平轴按虚线所指方向滑动该 t 分布, 并检测标在图中的 4 个均值. 分布图中没有哪个位置能使全部 4 个平均值可以想象为来自这一分布的、典型的随机选取的观测值. 这就意味着 4 个均值不是相等的. 这样一来, 这张草图就是方差分析结果的图解显示. 图中表明所有 4 个功率的水平 (160W, 180W, 200W, 220W) 产生的蚀刻速率均值与其他任一个都不同, 换句话说, $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$.

对于很多多重比较问题来说, 这一方法虽说粗糙但却是有效的. 不过, 还有更有效的方法, 现在就简要讨论一下这些方法.

3.5.4 对照法

很多多重比较法都用到**对照**的思想. 考虑例 3.1 中晶片蚀刻实验问题. 因为零假设被拒绝, 所以某些功率设置产生了不同于其他的蚀刻率, 但实际上是哪几种引起这一差异的呢? 可以推测, 在实验之初, 200 W 和 220 W 产生的蚀刻率相同, 也就是说, 我们要检验假设

$$H_0: \mu_3 = \mu_4, \qquad H_1: \mu_3 \neq \mu_4$$

或者

$$H_0: \mu_3 - \mu_4 = 0, \qquad H_1: \mu_3 - \mu_4 \neq 0 \tag{3.23}$$

实验初始阶段, 如果推测功率的低水平的均值与功率的高水平的均值相同, 则假设为

$$H_0: \mu_1 + \mu_2 = \mu_3 + \mu_4, \qquad H_1: \mu_1 + \mu_2 \neq \mu_3 + \mu_4$$

即

$$H_0: \mu_1 + \mu_2 - \mu_3 - \mu_4 = 0, \qquad H_1: \mu_1 + \mu_2 - \mu_3 - \mu_4 \neq 0 \tag{3.24}$$

一般说来, **对照** (contrast) 是形如

$$\Gamma = \sum_{i=1}^{a} c_i \mu_i$$

的参数线性组合. 其中**对照系数** $c_1, c_2, \cdots, c_a$ 的和为零, 即 $\sum_{i=1}^{a} c_i = 0$. 于是, 前面的假设都可以用对照形式表达为:

$$H_0: \sum_{i=1}^{a} c_i \mu_i = 0, \qquad H_1: \sum_{i=1}^{a} c_i \mu_i \neq 0 \tag{3.25}$$

(3.23) 式中的假设的对照系数为 $c_1 = c_2 = 0, c_3 = +1, c_4 = -1$; 而 (3.24) 式中的假设的对照系数为 $c_1 = c_2 = +1, c_3 = c_4 = -1$.

用两种基本方法检验对照的假设. 第一种方法用 t 检验. 用**处理均值**的形式写出我们感兴趣的对照:

$$C = \sum_{i=1}^{a} c_i \bar{y}_{i.}$$

C 的方差是

$$V(C) = \frac{\sigma^2}{n} \sum_{i=1}^{a} c_i^2 \tag{3.26}$$

这里每种处理的样本大小是相等的. 如果 (3.25) 式的零假设是真的, 则比

$$\frac{\sum_{i=1}^{a} c_i \bar{y}_{i.}}{\sqrt{\frac{\sigma^2}{n} \sum_{i=1}^{a} c_i^2}}$$

服从 $N(0,1)$ 分布. 现在用它的估计量 (均方误差 MS_E) 替代未知方差 σ^2, 用统计量

$$t_0 = \frac{\sum_{i=1}^{a} c_i \overline{y}_{i.}}{\sqrt{\frac{MS_E}{n}\sum_{i=1}^{a} c_i^2}} \tag{3.27}$$

检验 (3.25) 式中的假设. 如果 (3.27) 式中的 $|t_0|$ 超过 $t_{\alpha/2,n-a}$, 则零假设将被拒绝.

第二种方法用 F 检验. 自由度为 ν 的 t 随机变量的平方服从分子自由度为 1、分母自由度为 ν 的 F 分布. 因而, 我们用

$$F_0 = t_0^2 = \frac{\left(\sum_{i=1}^{a} c_i \overline{y}_{i.}\right)^2}{\frac{MS_E}{n}\sum_{i=1}^{a} c_i^2} \tag{3.28}$$

作为检验 (3.25) 式的 F 统计量. 当 $F_0 > F_{\alpha,1,N-a}$ 时, 拒绝零假设. 可以写出 (3.28) 式中的检验统计量为

$$F_0 = \frac{MS_C}{MS_E} = \frac{SS_C/1}{MS_E}$$

其中单个自由度的对照的平方和为

$$SS_C = \frac{\left(\sum_{i=1}^{a} c_i \overline{y}_{i.}\right)^2}{\frac{1}{n}\sum_{i=1}^{a} c_i^2} \tag{3.29}$$

1. 对照的置信区间

与检验关于对照的假设相比, 构建置信区间更为有用. 假设感兴趣的对照为

$$\Gamma = \sum_{i=1}^{a} c_i \mu_i$$

用处理的样本平均代替处理均值, 有

$$C = \sum_{i=1}^{a} c_i \overline{y}_{i.}$$

因为

$$\mathrm{E}\left(\sum_{i=1}^{a} c_i \overline{y}_{i.}\right) = \sum_{i=1}^{a} c_i \mu_i, \quad V(C) = \frac{\sigma^2}{n}\sum_{i=1}^{a} c_i^2$$

所以, 对照 $\sum_{i=1}^{a} c_i \mu_i$ 的 $100(1-\alpha)\%$的置信区间为

$$\sum_{i=1}^{a} c_i \overline{y}_{i.} - t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}\sum_{i=1}^{a} c_i^2} \leqslant \sum_{i=1}^{a} c_i \mu_i \leqslant \sum_{i=1}^{a} c_i \overline{y}_{i.} + t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}\sum_{i=1}^{a} c_i^2} \tag{3.30}$$

注意, 我们是用 MS_E 来估计 σ^2 的. 如果 (3.30) 式包含零, 那么很明显我们将不能拒绝 (3.25) 式中的零假设.

2. 标准化对照

当感兴趣的对照不止一个时,用相同的尺度进行度量通常是有用的. 这样做的方法是将对照标准化使得它有方差 σ^2. 如果对照 $\sum_{i=1}^{a} c_i\mu_i$ 用处理总和的形式写为 $\sum_{i=1}^{a} c_i\overline{y}_{i.}$, 则用 $\sqrt{(1/n)\sum_{i=1}^{a} c_i^2}$ 去除就得到方差为 σ^2 的标准化对照. 有效地, **标准化对照**为

$$\sum_{i=1}^{a} c_i^*\overline{y}_{i.}$$

其中

$$c_i^* = \frac{c_i}{\sqrt{\frac{1}{n}\sum_{i=1}^{a} c_i^2}}$$

3. 不等的样本容量

当每个处理的样本容量不等时, 我们需要对前面的结论稍作修改. 首先, 注意对照的定义现在要求

$$\sum_{i=1}^{a} n_i c_i = 0$$

其他必要的改动是直接的, 例如, (3.27) 式中的 t 统计量变成了

$$t_0 = \frac{\sum_{i=1}^{a} c_i\overline{y}_{i.}}{\sqrt{MS_E\sum_{i=1}^{a}\frac{c_i^2}{n_i}}}$$

(3.29) 式中的对照平方和为

$$SS_C = \frac{\left(\sum_{i=1}^{a} c_i\overline{y}_{i.}\right)^2}{\sum_{i=1}^{a}\frac{c_i^2}{n_i}}$$

3.5.5 正交对照法

3.5.4 节中一个有用的特殊情况是**正交对照法**. 系数为 $\{c_i\}$ 与 $\{d_i\}$ 的两个对照是正交的, 如果

$$\sum_{i=1}^{a} c_i d_i = 0$$

对于不平衡设计, 需要满足的条件则是

$$\sum_{i=1}^{a} n_i c_i d_i = 0$$

对于 a 个处理, $a-1$ 组正交对照将处理的平方和分解为 $a-1$ 个独立的单自由度分量. 这样一来, 在正交对照上进行的检验就是相互独立的.

对一组处理来说, 有多种方法来选择正交对照系数. 通常, 实验的某些本质方面应该提示我们, 哪些比较是我们感兴趣的. 例如, 如果有 a=3 个处理, 处理 1 作为一个控制, 处理 2 与处理 3 是实验者感兴趣的因子的实际水平, 则合适的正交对照如下:

处　理	正交对照的系数	
1(控制)	−2	0
2(水平 1)	1	−1
3(水平 2)	1	1

注意, $c_i = -2, 1, 1$ 的对照 1 对因子的平均效应与控制进行比较, 而 $d_i = 0, -1, 1$ 的对照 2 对我们感兴趣的因子的两个水平进行比较.

一般地, 对照 (或正交对照) 的方法对于所谓的**计划前对照**是有用的. 也就是, 对照系数必须在进行实验与检测数据之前选定. 这一点的理由是, 如果比较是在检测数据之后选定, 则大多数实验者构造出的检验法对应于均值可能会有大的观测差异. 这些大的差异可能是实际效应出现的结果, 也可能是随机误差的结果. 如果实验者总是挑最大的差异来比较, 他们就会增大检验的第 I 类错误, 因为, 以不寻常的高百分率来选定那些比较时, 观测得到的差异很可能就是误差的结果. 查看数据去选择感兴趣的对照通常被称为**数据探究**. 在 3.5.6 节讨论的适合所有对照的 Scheffé方法允许采用数据探究.

例 3.6 考虑例 3.1 中的晶片蚀刻问题. 有 4 个处理均值和 3 个处理间的自由度. 假定在实验前指定了以下的这些处理均值之间的几组比较 (和有关的对照) 是:

假　设	对　照
$H_0: \mu_1 = \mu_2$	$C_1 = \bar{y}_{1.} - \bar{y}_{2.}$
$H_0: \mu_1 + \mu_2 = \mu_3 + \mu_4$	$C_2 = \bar{y}_{1.} + \bar{y}_{2.} - \bar{y}_{3.} - \bar{y}_{4.}$
$H_0: \mu_3 = \mu_4$	$C_3 = \bar{y}_{3.} - \bar{y}_{4.}$

注意, 这些对照系数是正交的. 用表 3.4 的数据, 得出对照的数值与平方和如下:

$$C_1 = +1(551.2) - 1(587.4) = -36.2 \qquad SS_{C_1} = \frac{(-36.2)^2}{\frac{1}{5}(2)} = 3\,276.10$$

$$C_2 = \begin{array}{l} +1(551.2) + 1(587.4) \\ -1(625.4) - 1(707.0) \end{array} = -193.8 \qquad SS_{C_2} = \frac{(-193.8)^2}{\frac{1}{5}(4)} = 46\,948.05$$

$$C_3 = +1(625.4) - 1(707.6) = -81.6 \qquad SS_{C_3} = \frac{(-81.6)^2}{\frac{1}{5}(2)} = 16\,646.40$$

这些对照的平方和完全地剖分处理平方和. 关于这种正交对照的检验通常合并到方差分析中, 如表 3.11 所示. 我们从 P 值得出的结论为, 在 $\alpha = 0.05$ 水平, 功率设置的水平 1 与水平 2 以及水平 3 与水平 4 之间有显著性差异, 水平 1 与水平 2 的平均值和水平 3 与水平 4 的平均值之间有显著性差异.

表 3.11　晶片蚀刻实验的方差分析

方差来源	平方和	自由度	均　方	F_0	P 值
功率设置	66 870.55	3	22 290.18	66.80	< 0.001
正交对照					
$C_1: \mu_1 = \mu_2$	(3276.10)	1	3276.10	9.82	< 0.01
$C_2: \mu_1 + \mu_3 = \mu_3 + \mu_4$	(46 948.05)	1	46 948.05	140.69	< 0.001
$C_3: \mu_3 = \mu_4$	(16 646.40)	1	16 646.40	49.88	< 0.001
误差	5 339.20	16	333.70		
总和	72 209.75	19			

3.5.6 用来比较全部对照的 Scheffé 方法

在很多时候, 实验者预先并不知道他们想要比较哪些对照, 或者, 他们感兴趣于多于 $a-1$ 个可能的比较. 在很多探索性的实验中, 感兴趣的比较仅仅在进行了数据的初步检测之后才能

发现. Scheffé(1953) 提出一种方法, 可以用来比较处理均值之间的任意一个或所有可能的对照. 在 Scheffé方法中, 任一可能比较的第 I 类错误至多是 α.

假设我们已确定了感兴趣的处理均值中的一组 m 个对照组成的集合

$$\Gamma_u = c_{1u}\mu_1 + c_{2u}\mu_2 + \cdots + c_{au}\mu_a \quad u = 1, 2, \cdots, m \tag{3.31}$$

以处理平均值 $\overline{y}_{i.}$ 中对应的对照是

$$C_u = c_{1u}\overline{y}_{1.} + c_{2u}\overline{y}_{2.} + \cdots + c_{au}\overline{y}_{a.} \quad u = 1, 2, \cdots, m \tag{3.32}$$

且这一对照的**标准误** (standard error) 是

$$S_{C_u} = \sqrt{MS_E \sum_{i=1}^{a}(c_{iu}^2/n_i)} \tag{3.33}$$

其中 n_i 是第 i 个处理中观测值的个数. 可以证明, 用来比较 C_u 的临界值是

$$S_{\alpha,u} = S_{C_u}\sqrt{(a-1)F_{\alpha,a-1,N-a}} \tag{3.34}$$

为检验对照 Γ_u 与零有显著差异这一假设, 可参照 C_u 及其临界值. 当 $|C_u| > S_{\alpha,u}$ 时, 拒绝对照 Γ_u 等于零的假设.

Scheffé方法还可用来构造处理均值间的所有可能对照的置信区间. 所得的区间 (不妨假设为 $C_u - S_{\alpha,u} \leqslant \Gamma_u \leqslant C_u + S_{\alpha,u}$) 是**联合置信区间**(simultaneous confidence interval), 也就是它们同时成立的概率至少是 $1 - \alpha$.

为了说明这一方法, 考虑例 3.1 中的数据并设感兴趣的对照是

$$\Gamma_1 = \mu_1 + \mu_2 - \mu_3 - \mu_4, \qquad \Gamma_2 = \mu_1 - \mu_4$$

这些对照的数值是

$$C_1 = \overline{y}_{1.} + \overline{y}_{2.} - \overline{y}_{3.} - \overline{y}_{4.} = 551.2 + 587.4 - 625.4 - 707.0 = -193.80$$

$$C_2 = \overline{y}_{1.} - \overline{y}_{4.} = 551.2 - 707.0 = -155.8$$

从 (3.33) 式知, 相应的标准误分别为

$$S_{C_1} = \sqrt{MS_E \sum_{i=1}^{5}(c_{i1}^2/n_i)} = \sqrt{333.70(1+1+1+1)/5} = 16.34$$

$$S_{C_2} = \sqrt{MS_E \sum_{i=1}^{5}(c_{i2}^2/n_i)} = \sqrt{333.70(1+1)/5} = 11.55$$

由 (3.34) 式得, 1%临界值是

$$S_{0.01,1} = S_{C_1}\sqrt{(a-1)F_{0.01,a-1,N-a}} = 16.34\sqrt{3(5.29)} = 65.09$$

$$S_{0.01,2} = S_{C_2}\sqrt{(a-1)F_{0.01,a-1,N-a}} = 11.55\sqrt{3(5.29)} = 45.97$$

因为 $|C_1| > S_{0.01,1}$, 所以我们得出结论, 对照 $\Gamma_1 = \mu_1 + \mu_2 - \mu_3 - \mu_4$ 不等于零, 即功率设置 1 和设置 2 作为一组的蚀刻率的均值与功率设置 3 和设置 4 作为一组的均值存在差异. 更进一步, 因为 $|C_2| > S_{0.01,1}$, 我们得出结论, 对照 $\Gamma_2 = \mu_1 - \mu_4$ 不等于零, 即处理 1 与处理 4 的蚀刻率均值显著不同.

3.5.7 处理均值的配对比较法

许多实际情况中, 我们希望只比较**配对均值**. 我们常常通过检验所有配对处理均值间的差异来决定哪些均值不同. 因此, 我们感兴趣的是形如 $\Gamma = \mu_i - \mu_j$(对于一切 $i \neq j$) 的对照. 尽管前面介绍的 Scheffé方法很容易解决这样的问题, 但它不是对这类比较最灵敏的方法. 我们现在转而考虑特别构造的用于所有 a 个总体均值之间的配对比较的方法.

假定我们想要比较所有 a 个处理均值的配对, 那么, 想要检验的零假设就是 $H_0 : \mu_i = \mu_j$(对于一切 $i \neq j$). 有大量的解决这类问题的方法. 现给出用于这类比较的两种普遍方法.

1. Tukey 检验

假定在方差分析 F 检验法下拒绝了等处理均值的零假设后, 我们要对一切 $i \neq j$ 检验所有配对均值比较

$$H_0 : \mu_i = \mu_j, \qquad H_1 : \mu_i \neq \mu_j$$

Tukey(1953) 提出了使得当样本容量相等时总的显著性水平等于 α、样本容量不等时显著性水平至多是 α 的假设检验方法. 他的方法也可以用于所有配对均值比较的置信区间. 对于这些区间, 样本量相等时, 联合置信水平为 $100(1-\alpha)\%$; 而样本量不等时, 联合置信水平至少为 $100(1-\alpha)\%$. 换句话说, Tukey 方法将**实验总体**错误率即 “族” 错误率控制在选定的水平 α 上. 当着眼于配对均值时, 这是一个很好的数据探究方法.

Tukey 法利用**学生化极差统计量** (studentized range statistic)

$$q = \frac{\overline{y}_{\max} - \overline{y}_{\min}}{\sqrt{MS_E/n}}$$

的分布, 其中 $\overline{y}_{\max}$ 和 $\overline{y}_{\min}$ 分别是 p 个样本均值中的最大值和最小值. 附表Ⅶ中包含了 $q_\alpha(p,f)$(q 的上 $\alpha\%$点, 其中 f 是 MS_E 的自由度) 的值. 当样本容量相等时, 如果样本均值差的绝对值超过

$$T_\alpha = q_\alpha(a,f)\sqrt{\frac{MS_E}{n}} \tag{3.35}$$

Tukey 检验就表明两个均值有显著性差异. 等价地, 我们可以构建一系列关于所有配对均值的 $100(1-\alpha)\%$的置信区间如下:

$$\overline{y}_{i.} - \overline{y}_{j.} - q_\alpha(a,f)\sqrt{\frac{MS_E}{n}} \leqslant \mu_i - \mu_j \leqslant \overline{y}_{i.} - \overline{y}_{j.} + q_\alpha(a,f)\sqrt{\frac{MS_E}{n}}, \quad i \neq j. \tag{3.36}$$

当样本容量不等时, (3.35) 式和 (3.36) 式分别变为

$$T_\alpha = \frac{q_\alpha(a,f)}{\sqrt{2}}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \tag{3.37}$$

和

$$\begin{aligned}\overline{y}_{i.} - \overline{y}_{j.} - \frac{q_\alpha(a,f)}{\sqrt{2}}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} &\leqslant \mu_i - \mu_j \\ \leqslant \overline{y}_{i.} - \overline{y}_{j.} + \frac{q_\alpha(a,f)}{\sqrt{2}}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}, &\quad i \neq j\end{aligned} \tag{3.38}$$

样本量不等的方法有时也称为**Tukey-Kramer**法.

例 3.7 为了说明 Tukey 检验, 我们利用例 3.1 中的晶片蚀刻实验的数据. 当 $\alpha = 0.05$, 误差的自由度 f=16 时, 由附表Ⅶ查得 $q_{0.05}(4,16) = 4.05$, 因此, 由 (3.35) 式

$$T_{0.05} = q_{0.05}(4,16)\sqrt{\frac{MS_E}{n}} = 4.05\sqrt{\frac{333.70}{5}} = 33.09$$

所以, 当任意一对处理均值差异的绝对值超过 33.09 时, 这表示相应的配对总体有显著性差异. 4 个处理均值是

$$\overline{y}_{1.} = 551.2, \quad \overline{y}_{2.} = 587.4, \quad \overline{y}_{3.} = 625.4, \quad \overline{y}_{4.} = 707.0$$

且均值的差异为

$$\begin{aligned}
\overline{y}_{1.} - \overline{y}_{2.} &= 551.2 - 587.4 = -36.20^* \\
\overline{y}_{1.} - \overline{y}_{3.} &= 551.2 - 625.4 = -74.20^* \\
\overline{y}_{1.} - \overline{y}_{4.} &= 551.2 - 707.0 = -155.8^* \\
\overline{y}_{2.} - \overline{y}_{3.} &= 587.4 - 625.4 = -38.0^* \\
\overline{y}_{2.} - \overline{y}_{4.} &= 587.4 - 707.0 = -119.6^* \\
\overline{y}_{3.} - \overline{y}_{4.} &= 625.4 - 707.0 = -81.60^*
\end{aligned}$$

标有星号的值表示那对均值有显著性差异. 注意到 Tukey 方法表明所有配对均值都有显著差异. 即, 由每种功率设置得到的蚀刻率均值和由其他任一个功率设置得到的蚀刻率均值都有差异.

当使用均值配对检验的任一方法时, 我们有时会发现, 虽然方差分析中的总 F 统计量表明了某种显著性, 但均值的配对比较却找不到任一对有显著差异. 这种情况之所以会出现, 是因为 F 检验法同时考虑了处理均值之间的所有可能的比较, 而不仅仅是配对比较. 即, 在所掌握的数据中, 有显著性的对照不一定以 $\mu_i - \mu_j$ 的形式出现.

当样本容量相等时, (3.36) 式的 Tukey 置信区间的推导是直接的. 由学生化极差统计量 q, 我们有

$$\mathrm{P}\left(\frac{\max(\overline{y}_{i.} - \mu_i) - \min(\overline{y}_{i.} - \mu_i)}{\sqrt{MS_E/n}} \leqslant q_\alpha(a,f)\right) = 1 - \alpha$$

如果 $\max(\overline{y}_{i.} - \mu_i) - \min(\overline{y}_{i.} - \mu_i)$ 小于或等于 $q_\alpha(a,f)\sqrt{MS_E/n}$, 则 $|(\overline{y}_{i.} - \mu_i) - (\overline{y}_{j.} - \mu_j)| \leqslant q_\alpha(a,f)\sqrt{MS_E/n}$ 对每个均值配对都成立. 因而

$$\mathrm{P}\left(-q_\alpha(a,f)\sqrt{\frac{MS_E}{n}} \leqslant \overline{y}_{i.} - \overline{y}_{j.} - (\mu_i - \mu_j) \leqslant q_\alpha(a,f)\sqrt{\frac{MS_E}{n}}\right) = 1 - \alpha$$

将表达式重新整理, 可以得到一系列如 (3.38) 式所示的关于 $\mu_i - \mu_j$ 的 $100(1-\alpha)\%$的联合置信区间.

2. Fisher 最小显著性差异 (LSD) 法

比较全部配对均值的 Fisher 方法可以控制每个个体的配对比较的误差率 α, 但不能控制整个实验或族的误差率. 该方法使用 t 统计量检验 $H_0: \mu_i = \mu_j$,

$$t_0 = \frac{\overline{y}_{i.} - \overline{y}_{j.}}{\sqrt{MS_E\left(\dfrac{1}{n_i} + \dfrac{1}{n_j}\right)}} \tag{3.39}$$

假设备择假设是双边的, 则当 $|\overline{y}_{i.} - \overline{y}_{j.}| > t_{\alpha/2.N-a}\sqrt{MS_E(1/n_i + 1/n_j)}$ 时, 可表明均值对 μ_i 与 μ_j 有显著性差异. 量

$$\text{LSD} = t_{\alpha/2,N-a}\sqrt{MS_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \tag{3.40}$$

称为**最小显著性差异** (least significant difference). 如果设计是平衡的, 则 $n_1 = n_2 = \cdots = n_a = n$, 且

$$\text{LSD} = t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}} \tag{3.41}$$

使用 Fisher 的 LSD 法时, 只要简单地将观测到的每一对平均值的差与对应的 LSD 比较就行了. 当 $|\overline{y}_{i.} - \overline{y}_{j.}| > \text{LSD}$ 时, 结论是总体均值 μ_i 与 μ_j 有差异. 也可以使用 (3.39) 式中的 t 统计量.

例 3.8　为解释这一方法, 我们采用例 3.1 的数据. 当 $\alpha = 0.05$ 时, LSD 是

$$\text{LSD} = t_{0.025\ 16}\sqrt{\frac{2MS_E}{n}} = 2.120\sqrt{\frac{2(333.70)}{5}} = 24.49$$

这样一来, 任一对处理均值的差的绝对值大于 24.49 时, 就表示对应的一对总体均值有显著性差异. 均值的差是

$$\begin{aligned}
&\overline{y}_{1.} - \overline{y}_{2.} = 551.2 - 587.4 = -36.2^* &\quad& \overline{y}_{2.} - \overline{y}_{3.} = 587.4 - 625.4 = -38.0^*\\
&\overline{y}_{1.} - \overline{y}_{3.} = 551.2 - 625.4 = -74.2^* && \overline{y}_{2.} - \overline{y}_{4.} = 587.4 - 707.0 = -119.6^*\\
&\overline{y}_{1.} - \overline{y}_{4.} = 551.2 - 707.0 = -155.8^* && \overline{y}_{3.} - \overline{y}_{4.} = 625.4 - 707.0 = -81.6^*
\end{aligned}$$

标有星号的值表示那对均值有显著差异. 显然, 所有的配对比较都有显著差异.

注意, 使用这一方法, 可能会大大提高总体 α 的风险. 尤其是, 处理的数量 a 越大, 实验总体 (“族”) 的第 I 类错误就越大 (至少犯 1 个第 I 类错误的实验数与总的实验数的比值就越大).

3. *方法的选择*

当然, 此时一个自然的问题就是, 我应该用这些方法中的哪一种呢? 很遗憾, 这个问题并没有明显的一刀切的答案, 专业统计工作者常常不赞成利用不同的方法. Carmer and Swanson(1973) 引入 Monte Carlo 模拟, 研究了许多多重比较法, 包括在这里没有讨论的其他方法. 他们的报告指出, 如果方差分析的 F 检验仅在 5%显著时, 最小显著性差异法在检验真实的均值差方面是一种十分有效的检验法. 但是, 这种方法并不包含实验总体误差率. 因为 Tukey 方法确实能控制总的误差率, 许多统计工作者更喜欢使用它.

正如上面所指出的, 有很多其他的多重比较法. 描述这些方法的文献有 Miller (1977), O'Neill and Wetherill(1971), 以及 Nelson(1989). Miller(1966) 和 Hsu (1996) 的书亦有所论述.

3.5.8　将各个处理均值与一个控制进行比较

在很多实验中, 一种处理实际上就是一个**控制**, 实验分析就是要将其余 $a-1$ 种处理的每一种和这个控制进行比较. 这样一来, 就只有 $a-1$ 个比较要做. Dunnett(1964) 研究过进行这些比较的方法. 设处理 a 是控制, 则要对 $i = 1, 2, \cdots, a-1$ 检验假设

$$H_0 : \mu_i = \mu_a, \qquad H_1 : \mu_i \neq \mu_a$$

Dunnett 方法是普通的 t 检验法的一种修正. 对于每一种假设, 计算观测得到的样本均值差

$$|\overline{y}_{i.} - \overline{y}_{a.}| \quad i = 1, 2, \cdots, a-1$$

当

$$|\overline{y}_{i.} - \overline{y}_{a.}| > d_\alpha(a-1, f)\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_a}\right)} \tag{3.42}$$

时, 用第 I 类错误率 α 来拒绝零假设 $H_0: \mu_i = \mu_a$, 其中常数 $d_\alpha(a-1, f)$ 由附录表Ⅷ给出 (双边检验与单边检验都可能). 注意, α 是所有 $a-1$ 个检验的**联合显著性水平** (joint significance level).

例 3.9 用例 3.1 的数据来说明 Dunnett 检验法, 其中处理 4 作为控制. 在本例中, $a = 4, a - 1 = 3$, $f = 16$, $n_i = n = 5$. 取 5%水平, 从附录表 IX 查得 $d_{0.05}(3,16)$=2.59. 于是, 临界差为

$$d_{0.05}(3,16)\sqrt{\frac{2MS_E}{n}} = 2.59\sqrt{\frac{2(333.70)}{5}} = 29.92$$

[注意, 这是由平衡设计所得的 (3.42) 式的简式]. 这样一来, 任一处理均值与控制的绝对差大于 4.76 时, 就叫作有显著性差异. 这些观测值的差是

$$1 \text{ 对 } 4{:}\ \overline{y}_{1.} - \overline{y}_{4.} = 551.2 - 707.0 = -155.8$$

$$2 \text{ 对 } 4{:}\ \overline{y}_{2.} - \overline{y}_{4.} = 587.4 - 707.0 = -119.6$$

$$3 \text{ 对 } 4{:}\ \overline{y}_{3.} - \overline{y}_{4.} = 625.4 - 707.0 = -81.6$$

注意所有的差异都是显著的. 我们可以得到结论, 所有功率设置与控制都有差异.

当处理和一个控制比较时, 如果其余 $a-1$ 个处理的观测数相同, 则采用的控制处理的观测值 (记 n_a) 最好比其他处理的观测值 (记 n) 更多一些. 应该选择比值 n_a/n 近似等于处理总数的平方根. 即, 选择 $n_a/n = \sqrt{a}$.

3.6 计算机输出示例

有许多可用于实验设计和进行方差分析的计算机程序. 用 Design-Expert 处理例 3.1 的晶片蚀刻实验的数据, 得到的输出如图 3.12 所示. 输出中的 "Model"(模型) 的平方和是单因子设计的 $SS_{处理}$. 来源中的因子用 "A" 表示. 当实验中不只一个因子时, 模型平方和将被分成几个来源 (A, B, 等等). 计算机输出的顶部的方差分析表包括各平方和、自由度 (DF)、均方和检验统计量 F_0. 列 "Prob>F" 是 P 值 (事实上, 是 P 值的上界, 因为概率小于 0.000 1 被视为 0.000 1).

除了基本的方差分析之外, 程序显示了一些其他有用的信息. "R-Squared"(R 是复相关系数) 定义为

$$R^2 = \frac{SS_{模型}}{SS_{总}} = \frac{66\ 870.55}{72\ 209.75} = 0.926\ 1$$

大体上可解释为由方差分析模型所 "说明" 的数据中的变异性所占比例. 这样一来, 在晶片蚀刻实验中, "功率" 因子约占蚀刻速率变异性的 92.61%. 显然, $0 \leqslant R^2 \leqslant 1$ 恒成立, 且希望 R^2 有

ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	66870.55	3	22290.18	66.80	<0.0001 significant
A	*66870.55*	*3*	*22290.18*	*66.80*	*<0.0001*
Pure Error	5339.20	16	333.70		
Cor Total	72209.75	19			

The Model F-value of 66.80 implies the model is significant. There is only a 0.01% chance that a "Model F-Value" this large could occur due to noise.

Values of "Prob > F" less than 0.0500 indicate model terms are significant.
In this case A are significant model terms.
Values greater than 0.1000 indicate the model terms are not significant.
If there are many insignificant model terms (not counting those required to support hierarchy), model reduction may improve your model.

Std. Dev.	18.27	R-Squared	0.9261
Mean	617.75	Adj R-Squared	0.9122
C.V.	2.96	Pred R-Squared	0.8845
PRESS	8342.50	Adeq Precision	19.071

The "Pred R-Squared" of 0.8845 is in reasonable agreement with the "Adj R-Squared" of 0.9122.

"Adeq Precision" measures the signal to noise ratio. A ratio greater than 4 is disirable. Your ratio of 19.071 indicates an adequate signal. This model can be used to navigate the design space.

Treatment Means (Adjusted, If Necessary)

	Estimated Mean	Standard Error
1-160	551.20	8.17
2-180	587.40	8.17
3-200	625.40	8.17
4-220	707.00	8.17

Treatment	Mean Difference	DF	Standard Error	t for H_0 Coeff=0	Prob > |t|
1 vs 2	−36.20	1	11.55	−3.13	0.0064
1 vs 3	−74.20	1	11.55	−6.42	<0.0001
1 vs 4	−155.80	1	11.55	−13.49	<0.0001
2 vs 3	−38.00	1	11.55	−3.29	0.0046
2 vs 4	−119.60	1	11.55	−10.35	<0.0001
3 vs 4	−81.60	1	11.55	−7.06	<0.0001

Values of "Prob > |t|" less than 0.0500 indicate the difference in the treatment means is significant.
Values of "Prob > |t|" greater than 0.1000 indicate the difference in the two treatment means is not significant.

Diagnostics Case Statistics

Standard Order	Actual Value	Predicted Value	Residual	Leverage	Student Residual	Cook's Distance	Outlier t	Run Order
1	575.00	551.20	23.80	0.200	1.457	0.133	1.514	13
2	542.00	551.20	−9.20	0.200	−0.563	0.020	−0.551	14
3	530.00	551.20	−21.20	0.200	−1.298	0.105	−1.328	8
4	539.00	551.20	−12.20	0.200	−0.747	0.035	−0.736	5
5	570.00	551.20	18.80	0.200	1.151	0.083	1.163	4
6	565.00	587.40	−22.40	0.200	−1.371	0.117	−1.413	18
7	593.00	587.40	5.60	0.200	0.343	0.007	0.333	9
8	590.00	587.40	2.60	0.200	0.159	0.002	0.154	6
9	579.00	587.40	−8.40	0.200	−0.514	0.017	−0.502	16
10	610.00	587.40	22.60	0.200	1.383	0.120	1.427	17
11	600.00	625.40	−25.40	0.200	−1.555	0.151	−1.634	7
12	651.00	625.40	25.60	0.200	1.567	0.153	1.649	19
13	610.00	625.40	−15.40	0.200	−0.943	0.056	−0.939	10
14	637.00	625.40	11.60	0.200	0.710	0.032	0.699	20
15	629.00	625.40	3.60	0.200	0.220	0.003	0.214	1
16	725.00	707.00	18.00	0.200	1.102	0.076	1.110	2
17	700.00	707.00	−7.00	0.200	−0.428	0.011	−0.417	3
18	715.00	707.00	8.00	0.200	0.490	0.015	0.478	15
19	685.00	707.00	−22.00	0.200	−1.346	0.113	−1.385	11
20	710.00	707.00	3.00	0.200	0.184	0.002	0.178	12

Proceed to Diagnostic Plots (the next icon in progression). Be sure to look at the:
(1) Normal probability plot of the studentized residuals to check for normality of residuals.
(2) Studentized residuals versus predicted values to check for constant error.
(3) Outlier t versus run order to look for outliers, i.e., influential values.
(4) Box-Cox plot for power transformations.

If all the model statistics and diagnostic plots are OK, finish up with the Model Graphs icon.

图 3.12 例 3.1 的 Design-Expert 计算机输出

较大的值. 在输出中还有其他一些类似 R^2 的统计量. “调整后的”R^2 是对通常 R^2 统计量的调整, 它反映了模型的因子数. 对于有许多设计因子的更复杂的实验, 当我们要评估模型中项数增减产生的影响时, 它是一个有用的统计量. “Std Dev” 是误差均方的平方根, $\sqrt{333.70} = 18.27$, “C.V.” 是变异系数, 定义为 $(\sqrt{MS_E}/\overline{y})100$. 变异系数作为响应变量均值的百分率用来度量数据中未说明的或残余的变异性. “PRESS” 代表 “预期误差平方和”, 它度量的是用实验模型预测新实验中的响应的良好程度. PRESS 的值越小越好. 我们可以基于 PRESS 计算预测的 R^2(后面我们将说明如何做). 我们问题中的 R^2_{Pred} 是 0.884 5, 这是不合理的, 在当前实验中, 根据变异性的 93%来考虑模型. 用最大预测响应与最小预测响应的差除以所有预测响应的平均标准偏差, 计算 “Adeq Precision”(足够的精度) 统计量并这个量的值大一些是较为理想的, 超过 4 表明模型在预测方面有相当好的性能.

输出的结果中列出了处理均值的估计值以及标准误 (或每个处理均值的样本标准差, $\sqrt{MS_E/n}$). 输出结果中还通过利用 3.5.7 节中 Fisher 的 LSD 方法的假设检验结果显示了配对处理均值的差.

```
Source  DF     SS     MS      F      P
Power    3  66871  22290  66.80  0.000
Error   16   5339    334
Total   19  72210

S = 18.27   R-Sq = 92.61%   R-Sq (adj) = 91.22%

                                 Individual 95% CIs For Mean Based on
                                 Pooled StDev
Level  N    Mean  StDev    ---+---------+---------+---------+------
160    5  551.20  20.02    (--*---)
180    5  587.40  16.74           (---*---)
200    5  625.40  20.53                   (---*---)
220    5  707.00  15.25                                (---*---)
                           ---+---------+---------+---------+------
                            550       600       650       700

Pooled StDev = 18.27

Turkey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons among Levels of Power

Individual confidence level = 98.87%

Power = 160 subtracted from:

Power   Lower  Center   Upper  ----+---------+---------+---------+---
180      3.11   36.20   69.29            (---*---)
200     41.11   74.20  107.29                (---*----)
220    122.71  155.80  188.89                         (---*---)
                               ----+---------+---------+---------+---
                                -100         0       100       200

Power = 180 subtracted from:

Power   Lower  Center   Upper  ----+---------+---------+---------+---
200      4.91   38.00   71.09            (----*---)
220     86.51  119.60  152.69                     (---*---)
                               ----+---------+---------+---------+---
                                -100         0       100       200

Power = 200 subtracted from:

Power   Lower  Center   Upper  ----+---------+---------+---------+---
220     48.51   81.60  114.69                 (---*---)
                               ----+---------+---------+---------+---
                                -100         0       100       200
```

图 3.13 例 3.1 的 Minitab 计算机输出

计算机程序还计算并显示按 (3.16) 式定义的残差. 程序也生成了在 3.4 节中讨论的所有残差图. 输出中还显示了一些其他的残差诊断. 其中的部分内容将在以后讨论.

注意到, 计算机程序在输出中也有一些说明. 这类 "建议" 信息在许多 PC 机的统计软件包中的格式是统一的. 要记住, 它的写法是通用的, 也许不符合某个特定实验者的报告写作要求. 使用者也可以删去这些说明.

图 3.13 列出了 Minitab(另一种流行统计软件包) 关于晶片蚀刻实验的输出. 这个输出与图 3.12 的 Design-Expert 非常地类似. 它提供了每个单个处理均值的置信区间, 并用 Tukey 方法比较了配对均值. 但是, 所列出的 Tukey 方法使用的是置信区间形式, 而不是 3.5.7 节中的假设检验形式. Tukey 置信区间都不包括零, 所以, 我们得出所有均值都有显著差异的结论.

3.7 确定样本量

在任一实验设计问题中, 一项重大的决策就是样本量的选择, 即, 确定实验的重复次数. 一般来说, 如果要检测出较小的效应, 则与要检测出较大的效应相比, 实验者需要更多的重复次数. 本节讨论几种确定样本量的解决方法. 虽然我们的讨论着眼于单因子设计, 但大多数方法都可用于更为复杂的实验设计.

3.7.1 抽检特性曲线

抽检特性曲线是某一特定样本量的统计检验犯第 II 类错误的概率与某一反映零假设不成立程度的参数的关系图. 这些曲线可以用于指导实验者选择重复的次数, 使得设计对处理中的重要的潜在差异反应灵敏.

对于每种处理的样本量相等时的情形, 我们考虑固定效应模型的第 II 类错误的概率, 即

$$\beta = 1 - \mathrm{P}\{\text{拒绝 } H_0|H_0 \text{ 不真}\} = 1 - \mathrm{P}\{F_0 > F_{\alpha,a-1,N-a}|H_0 \text{ 不真}\} \tag{3.43}$$

要计算 (3.43) 式中的概率, 需要知道零假设不真时检验统计量 F_0 的分布. 可以证明, 当 H_0 不真时, 统计量 $F_0 = MS_{\text{处理}}/MS_E$ 的分布是自由度为 $a-1$ 与 $N-a$, 非中心参数为 δ 的**非中心F分布**. 当 $\delta = 0$ 时, 非中心分布就是通常的 (中心)F 分布.

附录的图 V 给出的抽检特性曲线可用来估算 (3.43) 式的概率. 这些曲线画出了第 II 类错误概率 (β) 与参数 Φ 的关系, 其中

$$\Phi^2 = \frac{n\sum_{i=1}^{a}\tau_i^2}{a\sigma^2} \tag{3.44}$$

量 Φ^2 与非中心参数 δ 有关. 这些曲线适用于 $\alpha = 0.05$ 和 $\alpha = 0.01$ 以及一定范围的分子自由度和分母自由度.

在使用 OC 曲线时, 实验者必须确定参数 Φ 和 σ^2 的值. 但在实际中做到这一点往往很难. 确定 Φ 的一种办法是, 选择我们想以高概率拒绝零假设的处理均值的实际数值. 于是, 当 $\mu_1, \mu_2, \cdots, \mu_a$ 是指定的处理均值时, 用 $\tau_i = \mu_i - \overline{\mu}$ 求得 τ_i, 其中 $\overline{\mu} = (1/a)\sum_{i=1}^{a}\mu_i$ 是各个处理均值的平均值. 我们还需要 σ^2 的估计量. 有时, 可以使用先验的或前次实验得到的 (如第 1 章

建议的), 或判断得到的估计量. 当我们不能肯定 σ^2 的值时, 可以先对 σ^2 可能取值的范围确定样本量, 在作出最后的选择之前, 先研究这个参数对所需的样本量的效应.

例 3.10 考虑例 3.1 晶片蚀刻实验问题. 当 4 个处理均值是

$$\mu_1 = 575, \quad \mu_2 = 600, \quad \mu_3 = 650, \quad \mu_4 = 675$$

时, 实验者打算至少以 0.90 的概率拒绝零假设. 他计划用 $\alpha = 0.01$. 此时, 因为 $\sum_{i=1}^{4} \mu_i = 2\ 500$, 所以有 $\overline{\mu} = (1/4)2\ 500 = 625$ 以及

$$\tau_1 = \mu_1 - \overline{\mu} = 575 - 625 = -50$$
$$\tau_2 = \mu_2 - \overline{\mu} = 600 - 625 = -25$$
$$\tau_3 = \mu_3 - \overline{\mu} = 650 - 625 = 25$$
$$\tau_4 = \mu_4 - \overline{\mu} = 675 - 625 = 50$$

于是, $\sum_{i=1}^{4} \tau_i^2 = 6\ 250$. 设实验者觉察到在功率的任一水平上蚀刻率的标准差都不会大于 $\sigma = 25\text{Å}/\min$. 则由 (3.44) 式, 有

$$\Phi^2 = \frac{n\sum_{i=1}^{4}\tau_i^2}{a\sigma^2} = \frac{n(6\ 250)}{4(25)^2} = 2.5n$$

对于 $a-1=4-1=3$ 以及 $N-a=a(n-1)=4(n-1)$ 个误差自由度和 $\alpha = 0.01$, 用特性曲线 (见附录图表 V) 对所需样本量作第一次猜想. 试用 $n=3$ 次重复. 这产生 $\Phi^2 = 2.5n = 2.5(3) = 7.54$, $\Phi = 2.74$, 以及 4(2)=8 个误差自由度. 因此, 由图表 V, 求得 $\beta \approx 0.25$. 因此, 检验的势大约为 $1-\beta = 1-0.25 = 0.75$, 小于所需的 0.90. 结论是, $n=3$ 次重复不满足要求. 用同样的方法继续进行, 得出下表:

n	Φ^2	Φ	$a(n-1)$	β	势 $(1-\beta)$
3	7.5	2.74	8	0.25	0.75
4	10.0	3.16	12	0.04	0.96
5	12.5	3.54	16	< 0.01	> 0.99

于是, 为了达到所要求的势, 4~5 次重复足够了.

利用抽检特性曲线方法, 有一个明显的问题是, 很难选择一组作为决定样本量的处理均值. 另一种选择样本量的方法是, 只要有两个处理均值之差超过一个规定值时就拒绝零假设. 如果任意两个处理均值之差都和 D 一样大, 则可证明 Φ^2 的最小值是

$$\Phi^2 = \frac{nD^2}{2a\sigma^2} \tag{3.45}$$

因为这是 Φ^2 的最小值, 所以由抽检特性曲线求得的对应样本量是一个保守的值. 也就是说, 它提供的势至少如实验者所规定的那样大.

为了说明这种方法, 假设在例 3.1 的晶片蚀刻实验问题中, 当任意两个处理均值的差不超过 $75\text{Å}/\min$ 时, 实验者想以概率至少为 0.90 来拒绝零假设. 于是, 假定 $\sigma = 25\text{Å}/\min$, 求得 Φ^2 的最小值是

$$\Phi^2 = \frac{n(75)^2}{2(4)(25^2)} = 1.125n$$

现在, 我们可以在例 3.10 中精确地使用 OC 曲线. 假定我们实验 $n=4$ 次重复. 结果是 $\Phi^2 = 1.125(4) = 4.5$, $\Phi = 2.12$, 误差的自由度 $4(3) = 12$. 从 OC 曲线可以发现势接近 0.65. 对于 $n=5$ 次重复, $\Phi^2 = 5.625, \Phi = 2.37$, 误差的自由度 $4(4) = 16$. 从 OC 曲线可以发现势接近 0.8. 对于 $n=6$ 次重复, $\Phi^2 = 6.75, \Phi = 2.60$, 误差的自由度 $4(5) = 20$. 从 OC 曲线可以发现势超过 0.90. 因此, 需要 $n=6$ 次重复.

Minitab 用这种方法计算势, 并找到单因子方差分析的样本容量. 如下图所示. 在表的上部我们要求 Minitab 计算 $n = 4$ 次重复的势. 注意这些结果非常接近 OC 曲线所示的结论. 表的下半部列出了为使目标势达到至少 0.90 的样本大小. 再一次, 结果与 OC 曲线的一致.

```
Power and Sample Size
One-way ANOVA
Alpha=0.01 Assumed standard deviation=25
Number of Levels=4
               Sample                      Maximum
SS Means         Size       Power       Difference
  2812.5            5    0.804838               75
The sample size is for each level.
Power and Sample Size
One-way ANOVA
Alpha=0.01 Assumed standard deviation=25
Number of Levels=4
            Sample   Target                          Maximum
SS Means      Size    Power    Actual Power       Difference
  2812.5         6      0.9        0.915384               75
The sample size is for each level.
```

3.7.2 规定标准差的增量

有时, 这种方法有助于确定样本量. 如果处理均值没有差异, 则随机选一个观测值, 其标准差是 σ. 如果处理均值是有差异的, 则随机选取的观测值的标准差是

$$\sqrt{\sigma^2 + \left(\sum_{i=1}^{a} \tau_i^2/a\right)}$$

如果为一个观测值的标准差的增量选取一个百分数 P, 超过该值就拒绝所有处理均值都相等的假设, 这等价于选择

$$\frac{\sqrt{\sigma^2 + \left(\sum_{i=1}^{a} \tau_i^2/a\right)}}{\sigma} = 1 + 0.01P \quad (P = \text{百分比})$$

即

$$\frac{\sqrt{\sum_{i=1}^{a} \tau_i^2/a}}{\sigma} = \sqrt{(1 + 0.01P)^2 - 1}$$

所以

$$\Phi = \frac{\sqrt{\sum_{i=1}^{a} \tau_i^2/a}}{\sigma/\sqrt{n}} = \sqrt{(1 + 0.01P)^2 - 1}(\sqrt{n}) \tag{3.46}$$

于是, 规定 P 的值, 就可由 (3.46) 式计算 Φ, 然后用附录中图 V 的抽检特性曲线确定所需要的样本量.

例如, 在例 3.1 的晶片蚀刻问题中, 设要以至少 0.90 的概率和 $\alpha = 0.05$ 检测出标准差有 20%的增量. 则

$$\Phi = \sqrt{(1.2)^2 - 1}(\sqrt{n}) = 0.66\sqrt{n}$$

查抽检特性曲线可知, 需要 n= 10 才能达到所希望的灵敏度.

3.7.3 置信区间的估计方法

此方法假定实验者想用置信区间表示最后的结果, 并且预先规定他们所要的置信区间的宽度. 例如, 设在例 3.1 的晶片蚀刻问题中, 任意两种功率设置的平均蚀刻速率之差是 ±30Å/ min, 求 95%置信区间, 而 σ 的先验估计量是 25. 于是, 用 (3.13) 式, 求得置信区间的准确度是

$$\pm t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}}$$

设试用 n=5 次重复. 用 $\sigma^2 = 25^2 = 625$ 作为 MS_E 的估计量, 则置信区间的准确度变为

$$\pm 2.120\sqrt{\frac{2(625)}{5}} = \pm 33.52$$

该结果达不到要求. 试用 n=6, 得

$$\pm 2.086\sqrt{\frac{2(625)}{6}} = \pm 30.11$$

试用 n=7 得

$$\pm 2.064\sqrt{\frac{2(625)}{7}} = \pm 27.58$$

显然, n=7 是达到所希望的准确度的最小样本量.

对引用的显著性水平, 上述说明仅适用于一个置信区间. 但是, 如果实验者想预先规定一组置信区间并对它们作出**联合置信区间**, 则可以用相同的一般的处理方案 (见 3.3.3 节中的联合置信区间的注释). 还有, 与上面说明的配对对照相比, 还可以为更为一般的处理均值的对照构造置信区间.

3.8 寻找分散效应

我们的着眼点在于, 利用方差分析以及相关的方法确定哪些因子水平会导致处理之间或因子水平均值之间的差异. 习惯上将这些效应看作为**位置效应** (location effect). 如果在不同的因子水平上方差不等, 则可以用方差稳定化变换来改善关于位置效应的推断. 但是, 在有些问题中, 我们想要发现不同的因子水平是否影响**变异性**. 也就是说, 我们想要寻找可能存在的**分散效应** (dispersion effect). 每当将标准差、方差或变异性的某些其他度量用作响应变量时, 就会出现这种情况.

为了说明这些思想, 考虑表 3.12 的数据, 它们是从炼铝厂的一个设计好的实验中得来的. 铝是通过在电解槽中将铝氧土和其他成分混合在一起并通过电流加热生产出来的. 铝氧土不断地

加进槽中, 使得槽中铝氧土与其他成分保持在合理比例水平上. 在此实验中, 研究了 4 种不同的比值控制法. 所研究的响应变量和电解槽的电压有关. 具体地说, 传感器每秒检测电压若干次, 在进行每一次实验时, 会得出上千个电压数据. 工程师决定用一次实验的平均电压和电压的标准差 (圆括号内的数) 作为响应变量. 平均电压是重要的, 因为它影响电解槽的温度; 电压的标准差 (工序工程师称之为 “电位噪音”) 也是重要的, 因为它影响整个电解槽的效率.

表 3.12　冶炼实验的数据

比值控制法	观测值					
	1	2	3	4	5	6
1	4.93(0.05)	4.86(0.04)	4.75(0.05)	4.95(0.06)	4.79(0.03)	4.88(0.05)
2	4.85(0.04)	4.91(0.02)	4.79(0.03)	4.85(0.05)	4.75(0.03)	4.85(0.02)
3	4.83(0.09)	4.88(0.13)	4.90(0.11)	4.75(0.15)	4.82(0.08)	4.90(0.12)
4	4.89(0.03)	4.77(0.04)	4.94(0.05)	4.86(0.05)	4.79(0.03)	4.76(0.02)

我们进行方差分析, 以确定不同的控制法是否影响电解槽的平均电压. 结果显示出比例控制法没有**位置效应**. 也就是说, 改变比值控制法并不改变电解槽的平均电压 (参见思考题 3.28).

为研究分散效应, 通常最好是用

$$\ln(s)\text{或}\ln(s^2)$$

作为响应变量, 因为对数变换对于样本标准差的分布能有效地稳定变异性. 当所有的电压的样本标准差都小于一个单位时, 就用

$$y=-\ln(s)$$

作为响应变量. 表 3.13 列出了这一响应 (即 “电位噪音” 的自然对数) 的方差分析. 比值控制法的选择对电位噪音有影响, 即比值控制法有**分散效应**. 标准的模型适合性检测法, 包括残差的正态概率图, 都表明实验的有效性是没有问题的 (参见思考题 35).

表 3.13　电位噪音的自然对数的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
比控制算法	6.166	3	2.055	21.96	< 0.001
误差	1.872	20	0.094		
总和	8.038	23			

图 3.14 画出了各种比值控制法的电位噪音的对数的平均值, 还介绍了尺度化的 t 分布图, 可用作识别比值控制法的**参考分布**. 这一图形清晰地揭示出, 比值控制法 3 比其他的方法产生较大的电位噪音或电解槽电压标准差. 而方法 1、方法 2 和方法 4 之间看来并没有很大的差别.

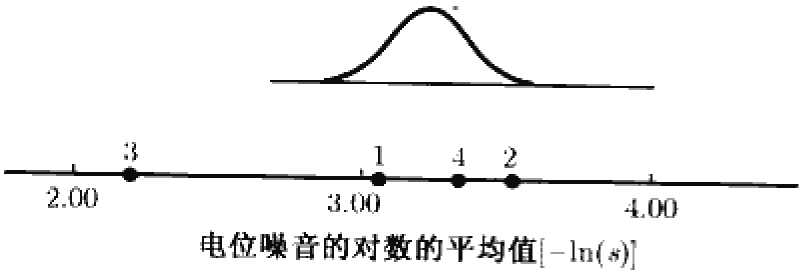


图 3.14　4 种比值控制法的电位噪音的对数的平均值 $[-\ln(s)]$ 与具有尺度因子 $\sqrt{MS_E/n}=\sqrt{0.094/6}=0.125$ 的尺度化 t 分布

3.9 方差分析的回归处理法

前面对方差分析进行了直观的或启发式的研究. 现在我们来进行更为正规的研究. 这种方法对于后面理解更复杂设计的统计分析基础将是有用的. 这种方法叫做**一般回归显著性检验法** (general regression significance test), 它的实质是为拟合模型与所包括的全体参数找出总平方和中的减少量, 以及当模型限制在零假设时的平方和中的减少量. 这两个平方和的差就是处理平方和, 利用它可导出一个对零假设的检验. 这一方法需要方差分析模型的参数的最小二乘估计量. 如在前面 3.3.3 节中所做的那样, 我们已经给出了这些参数估计. 但是, 现在我们要给出正规方法.

3.9.1 模型参数的最小二乘估计

现在, 我们用最小二乘法计算单因子模型

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

的参数的估计量. 为了计算 μ 和 τ_i 的最小二乘估计量, 首先写出误差的平方和公式

$$L = \sum_{i=1}^{a}\sum_{j=1}^{n}\epsilon_{ij}^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \mu - \tau_i)^2 \tag{3.47}$$

然后, 极小化 L, 选定 μ 和 τ_i 的估计, 即 $\hat{\mu}$ 和 $\hat{\tau}_i$. 合适的值是 $a+1$ 个联立方程的解:

$$\left.\frac{\partial L}{\partial \mu}\right|_{\hat{\mu},\hat{\tau}_i} = 0$$

$$\left.\frac{\partial L}{\partial \tau_i}\right|_{\hat{\mu},\hat{\tau}_i} = 0 \quad i = 1, 2, \cdots, a$$

(3.47) 式对 μ 和 τ_i 求微分, 得

$$-2\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \hat{\mu} - \hat{\tau}_i) = 0$$

和

$$-2\sum_{j=1}^{n}(y_{ij} + \hat{\mu} - \hat{\tau}_i) = 0 \quad i = 1, 2, \cdots, a$$

简化之, 得

$$\begin{aligned}
N\hat{\mu} + n\hat{\tau}_1 + n\hat{\tau}_2 + \cdots + n\hat{\tau}_a &= y_{..} \\
n\hat{\mu} + n\hat{\tau}_1 \qquad\qquad\qquad &= y_{1.} \\
n\hat{\mu} \qquad + n\hat{\tau}_2 \qquad\qquad &= y_{2.} \\
\vdots \qquad\qquad\qquad\qquad &\quad \vdots \\
n\hat{\mu} \qquad\qquad\qquad + n\hat{\tau}_a &= y_{a.}
\end{aligned} \tag{3.48}$$

有 $a+1$ 个未知数的 $a+1$ 个等式 [(3.48) 式] 叫做**最小二乘正规方程** (least square normal equation). 如果将后面 a 个正规方程相加, 就可以得到第一个正规方程. 因此, 正规方程不是线性无关的, 并没有 $\mu, \tau_1, \cdots, \tau_a$ 的唯一解. 但可以用一些方法克服这个困难. 因为我们已经定义处理效应为总体均值的偏差, 那么应用**约束**

$$\sum_{i=1}^{a} \hat{\tau}_i = 0 \tag{3.49}$$

似乎是合理的. 利用这个约束, 我们得到正规方程方程的解

$$\begin{aligned} \hat{\mu} &= \overline{y}_{..} \\ \hat{\tau}_i &= \overline{y}_{i.} - \overline{y}_{..}, \qquad i = 1, 2, \cdots, a \end{aligned} \tag{3.50}$$

显然, 这个解不是唯一的, 它取决于我们选择的约束 [(3.49) 式]. 这似乎是令人遗憾的, 因为如果两个不同的实验者用不同的约束分析相同的数据, 则可能得到不同的结论. 但是, 模型参数的函数是唯一估计的, 若不考虑约束的话. 例如, $\tau_i - \tau_j$ 一定是由 $\hat{\tau}_i - \hat{\tau}_j = \overline{y}_{i.} - \overline{y}_{j.}$ 估计, 第 i 个处理均值 $\mu_i = \mu + \tau_i$ 一定由 $\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \overline{y}_{i.}$ 估计.

因为我们通常感兴趣于处理效应间的差异, 而不是它们的真实值, 这使得 τ_i 不能被唯一估计无关要紧. 一般地, 若模型参数的任一函数是正规方程 [(3.48) 式] 左端项的线性组合, 则它可以被唯一估计. 不考虑所用约束的被唯一估计的函数叫做**估计函数**. 详见本章的**补充材料**. 现在, 我们准备在方差分析的一般研究中利用这些参数估计.

3.9.2 一般回归显著性检验

这一方法需要首先写出模型的正规方程组. 如在 3.9.1 节中所做的那样, 正规方程组通常是先构造最小二乘函数, 然后将此函数对各个未知参数微分而求得的. 不过, 还有一个更加容易的方法. 下面列出的法则就可对任一实验设计模型直接写出它的正规方程组.

法则 1 模型中每个待估计的参数总存在一个正规方程.

法则 2 任一正规方程的右端项恰是包含与这个指定的正规方程有关的参数的所有观测值之和.

为说明本法则, 考虑单因子模型. 第 1 个正规方程是关于参数 μ 的, 因此, 右端项是 $y_{..}$, 因为所有的观测值都包含 μ.

法则 3 任一正规方程的左端项是所有模型参数的和, 其中每个参数乘以它在右端项中出现的总次数. 参数加上尖号 (ˆ) 标志, 表示它们是**估计量**而不是真正的参数值.

例如, 考虑单因子实验的第一个正规方程, 按照上述法则, 应该是

$$N\hat{\mu} + n\hat{\tau}_1 + n\hat{\tau}_2 + \cdots + n\hat{\tau}_a = y_{..}$$

因为 μ 出现在所有的 N 个观测值中, τ_1 仅出现在第 1 种处理下所取的 n 个观测值中, τ_2 仅出现在第 2 种处理下所取的 n 个观测值中, 如此等等. 由 (3.48) 式, 我们证明了上述方程是正确的, 对应于 τ_1 的第 2 个正规方程是

$$n\hat{\mu} + n\hat{\tau}_1 = y_{1.}$$

因为只有第 1 种处理的观测值包含 τ_1 (这给出了右端项的 y_1), μ 和 τ_1 在 $y_{1.}$ 中恰好出现 n 次, 而所有其他的 τ_i 出现零次. 一般说来, 任一正规方程的左端项是右端项的期望值.

现在, 考虑找出由于给数据拟合一个指定模型所带来的在平方和中的减少量. 由于给数据拟合了一个模型, 我们也就 "解释了" 一些变异性. 也就是说, 我们给未解释的变异性减少一些量值. 未解释的变异性的减少量通常是参数估计量乘以对应于那个参数的正规方程的右端项之和. 例如, 在单因子实验中, 由于拟合**全模型**$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ 的减少量是

$$R(\mu,\tau)=\hat{\mu}y_{..}+\hat{\tau}_1y_{1.}+\hat{\tau}_2y_{2.}+\cdots+\hat{\tau}_ay_{a.}=\hat{\mu}y_{..}+\sum_{i=1}^{a}\hat{\tau}_iy_{i.} \tag{3.51}$$

记号 $R(\mu,\tau)$ 表示由于拟合包含 μ 和 $\{\tau_i\}$ 的模型而在平方和中的减少量. $R(\mu,\tau)$ 有时也叫做模型 $y_{ij}=\mu+\tau_i+\varepsilon_{ij}$ 的 "回归" 平方和. 与平方和中的减少量 (例如 $R(\mu,\tau)$) 有关的自由度通常等于线性无关的正规方程组的个数. 模型未顾及到的剩余的变异性由

$$SS_E=\sum_{i=1}^{a}\sum_{j=1}^{n}y_{ij}^2-R(\mu,\tau) \tag{3.52}$$

求得. 此量用作 $H_0:\tau_1=\tau_2=\cdots=\tau_a=0$ 的检验统计量的分母.

现在说明单因子实验的一般回归显著性检验法, 并表明它产生通常的单向方差分析. 模型是 $y_{ij}=\mu+\tau_i+\varepsilon_{ij}$, 由上述法则求得的正规方程组是

$$\begin{aligned}
N\hat{\mu}+n\hat{\tau}_1+n\hat{\tau}_2+\cdots+n\hat{\tau}_a &= y_{..}\\
n\hat{\mu}+n\hat{\tau}_1 &= y_{1.}\\
n\hat{\mu}\qquad+n\hat{\tau}_2 &= y_{2.}\\
\vdots\qquad\qquad &\ \ \vdots\\
n\hat{\mu}\qquad\qquad+n\hat{\tau}_a &= y_{a.}
\end{aligned}$$

将这些正规方程与 (3.48) 式比较一下.

应用约束 $\sum_{i=1}^{a}\hat{\tau}_i=0$, 得 μ 和 τ_i 的估计量是

$$\hat{\mu}=\overline{y}_{..}\quad \hat{\tau}_i=\overline{y}_{i.}-\overline{y}_{..}\quad i=1,2,\cdots,a$$

由拟合此全模型而在平方和中的减少量可由 (3.51) 式求得:

$$\begin{aligned}
R(\mu,\tau)&=\hat{\mu}y_{..}+\sum_{i=1}^{a}\hat{\tau}_iy_{i.}=(\overline{y}_{..})y_{..}+\sum_{i=1}^{a}(\overline{y}_{i.}-\overline{y}_{..})y_{i.}\\
&=\frac{y_{..}^2}{N}+\sum_{i=1}^{a}\overline{y}_{i.}y_{i.}-\overline{y}_{..}\sum_{i=1}^{a}y_{i.}=\sum_{i=1}^{a}\frac{y_{i.}^2}{n}
\end{aligned}$$

它有 a 个自由度, 因为有 a 个线性无关的正规方程. 由 (3.52) 式, 得误差平方和为

$$SS_E=\sum_{i=1}^{a}\sum_{j=1}^{n}y_{ij}^2-R(\mu,\tau)=\sum_{i=1}^{a}\sum_{j=1}^{n}y_{ij}^2-\sum_{i=1}^{a}\frac{y_{i.}^2}{n}$$

它有 $N-a$ 个自由度.

为求出由处理效应 ($\{\tau_i\}$) 所得的平方和, 我们考虑一个简化模型, 即, 受零假设 (即 $\tau_i=0$, 对一切 i) 约束的模型. 所得的简化模型是 $y_{ij}=\mu+\varepsilon_{ij}$. 此模型只有一个正规方程:

$$N\hat{\mu}=y_{..}$$

从而 μ 的估计量是 $\hat{\mu}=\overline{y}_{..}$. 于是, 由于拟合仅包含 μ 的模型所得的平方和的减少量是

$$R(\mu)=(\overline{y}_{..})(y_{..})=\frac{y_{..}^2}{N}$$

因为此简化模型仅有一个正规方程, 所以 $R(\mu)$ 有 1 个自由度. 若 μ 已在模型中, 则由 $\{\tau_i\}$ 贡献的平方和是 $R(\mu,\tau)$ 和 $R(\mu)$ 的差, 即

$$R(\tau|\mu) = R(\mu, \tau) - R(\mu) = R(\text{全模型}) - R(\text{简化模型}) = \frac{1}{n}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N}$$

它有 $a-1$ 个自由度. 我们已由 (3.9) 式知, 它就是 $SS_{\text{处理}}$. 作通常的正态性假定, 检验 $H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$ 的恰当的统计量是

$$F_0 = \frac{R(\tau|\mu)(/(a-1)}{\left[\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - R(\mu,\tau)\right] \Big/ (N-a)}$$

在零假设下, 它服从 $F_{a-1,N-a}$ 分布. 当然, 这就是单因子方差分析的检验统计量.

3.10 方差分析中的非参数方法

3.10.1 Kruskal-Wallis 检验法

当正态性假定不能被证明的情况下, 实验者希望有不依赖于正态性假定的检验法来代替方差分析 F 检验法. Kruskal and Wallis(1952) 提出了这样的一种方法. Kruskal-Wallis 检验法用来检验 a 种处理是完全一样的零假设, 而备择假设是由某些处理产生的观测值大于其他处理产生的观测值. 因为此方法设计需要对检验均值差很灵敏, 所以, 为方便起见, 将 Kruskal-Wallis 检验法视为检验处理均值相等的检验. 相对于通常的方差分析而言, Kruskal-Wallis 检验法是一种**非参数的方法**.

运用 Kruskal-Wallis 检验法时, 首先将观测值 y_{ij} 按升序排列, 然后将每一观测值用它的秩 (即位次, 比方说 R_{ij}) 来代替, 最小的观测值的秩是 1. 对有结 (观测值取值相同) 的情况, 则用它们的平均秩. 令 $R_{i.}$ 为第 i 种处理的秩和. 检验统计量是

$$H = \frac{1}{S^2}\left[\sum_{i=1}^{a}\frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}\right] \tag{3.53}$$

其中 n_i 是第 i 种处理的观测值的个数, N 是观测值的总个数, 以及

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{a}\sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4}\right] \tag{3.54}$$

注意到 S^2 正好是秩的方差. 如果没有结, 则 $S^2 = N(N+1)/12$, 且检验统计量简化为

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{a}\frac{R_{i.}^2}{n_i} - 3(N+1) \tag{3.55}$$

当结的个数是中等程度时, (3.54) 式和 (3.55) 式几乎没有差别, 且可以用较简单的形式 [(3.55) 式]. 如果 n_i 适当大, 比方说 $n_i \geqslant 5$, 则在零假设成立的条件下, H 近似服从 χ_{a-1}^2 分布. 因此, 如果

$$H > \chi_{\alpha,a-1}^2$$

则拒绝零假设. 也可以采用 P 值方法.

例 3.11 例 3.1 的数据及其对应的秩如表 3.14 所示. 因为有结, 所以我们用 (3.53) 式作为检验统计量. 由 (3.54) 式, 求得

$$S^2 = \frac{1}{19}\left[2\,869.50 - \frac{20(21)^2}{4}\right] = 34.97$$

检验统计量是

$$H = \frac{1}{S^2}\left[\sum_{i=1}^{a}\frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}\right] = \frac{1}{34.97}[2\ 796.30 - 2\ 205] = 16.91$$

因为 $H > \chi^2_{0.01,3} = 11.34$, 所以我们拒绝零假设, 结论是处理存在差异 ($H = 16.91$ 的 P 值 $= 7.38\times10^{-4}$). 这与普通方差分析 F 检验法所得的结论相同.

表 3.14 例 3.1 的晶片蚀刻实验的数据及其秩

功率							
160		180		200		220	
y_{1j}	R_{1j}	y_{2j}	R_{2j}	y_{3j}	R_{3j}	y_{4j}	R_{4j}
575	6	565	4	600	10	725	20
542	3	593	9	651	15	700	17
530	1	590	8	610	11.5	715	19
539	2	579	7	637	14	685	16
570	5	610	11.5	629	13	710	18
$R_{i.}$	17		39.5		63.5		90

3.10.2 关于秩变换的一般评论

3.10.1 节用观测值的秩替换观测值的方法叫做**秩变换** (rank transformation). 这是一种功效高, 应用广的方法. 如果我们将通常的 F 检验法应用于秩而不是应用于原始数据的话, 则得到

$$F_0 = \frac{H/(a-1)}{(N-1-H)/(N-a)} \tag{3.56}$$

作为检验统计量 [见 Conover(1980), p.337]. 当 Kruskal-Wallis 统计量 H 递增或递减时, F_0 也递增或递减, 所以, Kruskal-Wallis 检验法等价于将通常的方差分析应用于秩.

对于那些不存在相应于方差分析的非参数方法的实验设计问题, 秩变换仍有着广泛的适用性. 这包括了本书随后各章的大多数设计. 如果将数据进行秩变换并使用通常的 F 检验法, 则会得到有优良统计性质的近似方法 [请参阅 Conover and Iman (1976, 1981)], 当我们关心关于正态性假定或离群值的效应时, 建议对原始数据和秩都施行通常的方差分析. 当两种方法得出相似的结果时, 方差分析的假定条件可能会得到相当好的满足, 从而说明标准的分析法是令人满意的. 当两种方法的结果不同时, 则优先采用秩变换, 因为它可能较少受非正态性或不寻常的观测的干扰. 在这种情况下, 实验者可能要去研究使用非正态性变换, 检查数据和实验方法, 以便去确定是否出现离群值以及为什么会出现离群值.

3.11 思 考 题

*3.1 研究硅酸盐水泥砂浆的粘结抗拉强度. 可以经济地使用 4 种不同的混合方法. 做了一个完全随机化实验, 收集的数据如下:

混合方法	粘结抗拉 (lb/in^2)			
1	3 129	3 000	2 865	2 890
2	3 200	3 300	2 975	3 150
3	2 800	2 900	2 985	3 050
4	2 600	2 700	2 600	2 765

(a) 用 $\alpha = 0.05$ 检验混合方法影响水泥强度的假定.
(b) 用 3.5.3 节所讨论的图解法比较 4 种混合方法的粘结抗拉强度的均值. 你的结论是什么?
(c) 用 Fisher 的 LSD 方法作配对均值间的比较. $\alpha = 0.05$.
(d) 作残差的正态概率图. 关于正态假设的有效性, 结论是什么?
(e) 作残差与粘结抗拉强度预测值的关系图. 说明该图形.
(f) 作实验结果的散点图, 以解释实验的结论.

3.2 (a) 用 Tukey 检验再研究思考题 3.1 中的 (b), $\alpha = 0.05$. 由 Tukey 检验是否能得到与图解法和/或 Fisher 的 LSD 方法一样的结论?
(b) 解释 Tukey 和 Fisher 方法之间的差别.

*3.3 再考虑思考题 3.1 中的实验. 计算由 4 种混合方法中的每一种生产的硅酸盐水泥沙浆的粘结抗拉强度均值的 95%置信区间. 再计算方法 1 与方法 3 的均值差的 95%置信区间. 这有助于解释实验的结论吗?

*3.4 工程师正开发一个用于制造男士衬衫的新合成纤维. 通常拉伸强度受棉在混纺纤维中所占百分率的影响. 工程师用 5 种水平的棉含量做了完全随机化实验并重复实验 5 次. 数据如下表所示.

棉重百分率	观测值				
15	7	7	15	11	9
20	12	17	12	18	18
25	14	19	19	18	18
30	19	25	22	19	23
35	7	10	11	15	11

(a) 有证据支持棉含量影响拉伸强度均值的结论吗? 用 $\alpha = 0.05$.
(b) 用 Fisher 的 LSD 方法作配对均值间的比较. 你能得出什么结论?
(c) 分析该实验的残差, 阐明模型的适合性.

3.5 再考虑思考题 3.4 中的实验. 假设 30%的棉含量是一个控制. 用 Dunnett 检验将所有其他的均值与控制进行比较. 用 $\alpha = 0.05$.

*3.6 一位药品制造商想研究一种新药的生物活性. 用 3 种剂量水平做了一个完全随机单因子试验, 结果如下表所示.

剂量	观测值			
20 g	24	28	37	30
30 g	37	44	31	35
40 g	42	47	52	38

(a) 有证据支持剂量水平影响生物活性的结论吗? 用 $\alpha = 0.05$.
(b) 如果合适的话, 作配对均值间的比较. 你能得出什么结论?
(c) 分析该实验的残差, 阐明模型的适合性.

3.7 汽车租赁公司想研究被租赁汽车的类型是否影响租期的长短. 在一个特定租车点做了为期一周的实验. 为每一种汽车随机地选择了 10 个租赁合同. 结果如下表所示.

汽车类型	观测值									
超小汽车	3	5	3	7	6	5	3	2	1	6
小型汽车	1	3	4	7	5	6	3	2	1	7
中型汽车	4	1	3	5	7	1	2	4	2	7
大型汽车	3	5	7	5	10	3	4	7	2	7

(a) 有证据支持所租赁汽车的类型影响租赁合同的长短的结论吗? 用 $\alpha = 0.05$. 如果这样, 哪种类型的汽车对这个差异负责?

(b) 分析该实验的残差, 阐明模型的适合性.

(c) 注意到该实验的响应变量为计数型. 这是否会引发方差分析变异性的潜在问题?

*3.8 我是社区高尔夫俱乐部的成员. 我将一年分为 3 个高尔夫季节: 夏季 (6 月 ~9 月), 冬季 (11 月 ~3 月), 春秋季 (10 月, 4 月和 5 月). 我认为最好在夏季 (因为我有更多的时间且球场不太拥挤) 和春秋季 (因为球场不太拥挤) 打高尔夫. 在冬季最差 (因为很多临时住户都出现了. 球场拥挤, 打得慢, 我心情不佳). 去年的数据列在下表中.

季节	观测值									
夏季	83	85	85	87	90	88	88	84	91	90
春秋季	91	87	84	87	85	86	83			
冬季	94	91	87	85	87	91	92	86		

(a) 这些数据表明我的建议是正确的吗? 用 $\alpha = 0.05$.

(b) 分析该实验的残差, 阐明模型的适合性.

3.9 某地区歌剧团试图用 3 种方式从 24 个潜在赞助商获得捐款. 24 个潜在赞助商被分成 3 组, 每组 8 个, 一组用一种方法. 捐献结果以美元为单位列在下表中.

方法	捐款 (美元)							
1	1 000	1 500	1 200	1 800	1 600	1 100	1 000	1 250
2	1 500	1 800	2 000	1 200	2 000	1 700	1 800	1 900
3	900	1 000	1 200	1 500	1 200	1 550	1 000	1 100

(a) 这些数据表明 3 种不同方法得到的结果有差异吗? 用 $\alpha = 0.05$.

(b) 分析该实验的残差, 阐明模型的适合性.

3.10 进行一项实验, 确认 4 种不同的燃烧温度是否影响某种砖的密度. 完全随机化实验得到下面的数据:

温度	密度				
100	21.8	21.9	21.7	21.6	21.7
125	21.7	21.4	21.5	21.4	
150	21.9	21.8	21.8	21.6	21.5
175	21.9	21.7	21.8	21.4	

(a) 燃烧温度影响砖的密度吗? 用 $\alpha = 0.05$.

(b) 用 Fisher 的 LSD 方法比较均值合适吗?

(c) 分析实验的残差, 满足方差分析的假定吗?

(d) 作 3.5.3 节所讨论的处理图. 该图是否足以概括 (a) 的方差分析的结果?

3.11 用 Tukey 方法再研究思考题 3.10 中的 (d). 你得出什么样的结论? 详细解释你是如何修改该方法以适应样本大小不等的情形的.

*3.12 电视机厂感兴趣于彩色显像管 4 种不同的涂层对显像管的传导率是否有影响. 做完全随机化实验, 测得传导率的数据如下:

涂层类型	传导率			
1	143	141	150	146
2	152	149	137	143
3	134	136	132	127
4	129	127	132	129

(a) 涂层类型使传导率有差异吗? 用 $\alpha = 0.05$.

(b) 估计总均值与处理效应.

(c) 计算涂层 4 的均值的 95%置信区间估计. 计算涂层 1 与涂层 4 之间的均值差的 99%置信区间估计.

(d) 用 Fisher 的 LSD 方法检验所有的均值对, 用 $\alpha = 0.05$.

(e) 用 3.5.3 节所讨论的图解法对均值进行比较. 哪一类涂层得出最高的传导率?

(f) 假定现在用的是涂层 4, 想要使传导率最小, 你会向工厂推荐哪种涂层呢?

*3.13 再考虑思考题 3.12 的实验. 分析残差并得出关于模型适合性的结论.

3.14 *ACI Materials Journal* (Vol.84, 1987, pp. 213–216) 的一篇文章描述了几个实验, 研究混凝土的管道通条去除遗留的空气. 用一个 3 英寸 ×6 英寸的圆筒, 这种管道通条被使用的次数作为设计变量. 混凝土试样所导致的压强是响应. 数据如下表所示:

管道通条的水平	压强		
10	1 530	1 530	1 440
15	1 610	1 650	1 500
20	1 560	1 730	1 530
25	1 500	1 490	1 510

(a) 由管道通条的不同水平得到的压强有差异吗? 用 $\alpha = 0.05$.

(b) 计算 (a) 中 F 统计量的 P 值.

(c) 分析该实验的残差, 关于模型的基本假定你能得出什么结论?

(d) 构造一种图形显示, 来比较 3.5.3 节所讨论的处理均值.

*3.15 *Environment International* (Vol. 18, No.4, 1992) 中的一篇文章描述了一个研究喷水器释放氡的实验. 实验中用富含氡的水, 检验喷头的 6 种孔径, 实验得到下面的数据:

孔径	氡的释放率 (%)			
0.37	80	83	83	85
0.51	75	75	79	79
0.71	74	73	76	77
1.02	67	72	74	74
1.40	62	62	67	69
1.99	60	61	64	66

(a) 孔径大小影响氡的释放百分比吗? 用 $\alpha = 0.05$.

(b) 计算 (a) 中 F 统计量的 P 值.

(c) 分析该实验的残差.

(d) 当孔径为 1.40 时, 计算氡的释放率的均值的 95%置信区间.

(e) 用 3.5.3 节所讨论的图解法比较处理均值. 你能得出什么样的结论?

*3.16 用于自动断路开关装置中的 3 种不同类型的电路, 以毫秒为单位响应时间已被实验确定. 随机化实验的结果如下表所示:

电路类型	响应时间				
1	9	12	10	8	15
2	20	21	23	17	30
3	6	5	8	16	7

(a) 用 $\alpha = 0.01$ 检验 3 种电路类型有相同响应时间的假设.

(b) 用 Tukey 检验作配对处理均值间的比较. $\alpha = 0.01$.

(c) 构造 3.5.3 节所讨论的比较处理均值的图示方法. 你的结论是什么? 怎样将它们与从 (b) 得到的结论相比较?

(d) 构造一组正交对照, 假定在实验开始, 你怀疑第 2 类电路类型的响应时间和其他两种不同.

(e) 如果你是设计工程师, 希望使响应时间最短, 你将选择哪种电路类型?

(f) 分析该实验的残差. 满足方差分析的基本假定吗?

*3.17 研究在加速负荷为 35 千伏下的绝缘流体的有效寿命. 从 4 种类型的流体得到了检验数据. 完全随机化实验的结果如下:

流体类型	35 千伏下的寿命 (小时)						流体类型	35 千伏下的寿命 (小时)					
1	17.6	18.9	16.3	17.4	20.1	21.6	3	21.4	23.6	19.4	18.5	20.5	22.3
2	16.9	15.3	18.6	17.1	19.5	20.3	4	19.3	21.1	16.9	17.5	18.3	19.8

(a) 流体间有无显著性差异? 用 $\alpha = 0.05$.

(b) 假定希望目标是获得较长的寿命, 你将选择哪种流体?

(c) 分析该实验的残差. 满足方差分析的基本假定吗?

3.18 为了比较出现的噪音量, 研究数字计算机电路的 4 种不同的设计. 得到的数据如下:

电路设计	噪音观测量					电路设计	噪音观测量				
1	19	20	19	30	8	3	47	26	25	35	50
2	80	61	73	56	80	4	95	46	83	78	97

(a) 所有 4 种设计产生的噪音量相同吗? 用 $\alpha = 0.05$.

(b) 分析该实验的残差. 满足方差分析的基本假定吗?

(c) 依据噪音越低越好的原则, 你会选用哪种电路设计呢?

3.19 4 位化学家共同确定某一化学化合物中甲醇的百分含量. 每位化学家做了 3 次测试, 得出结果如下:

化学家	甲醇百分率			化学家	甲醇百分率		
1	84.99	84.04	84.38	3	84.72	84.48	85.16
2	85.15	85.13	84.88	4	84.20	84.10	84.55

(a) 化学家们的测试有显著性差异吗? 用 $\alpha = 0.05$.

(b) 分析该实验的残差.

(c) 如果化学家 2 是一位新的雇员, 试在实验前构造一组有意义的正交对照.

*3.20 研究 3 种商标的电池. 推测 3 种商标的电池寿命 (以周计) 是不同的, 每种商标随机检测 5 节电池, 得下述结果:

寿命 (周)		
商标 1	商标 2	商标 3
100	76	108
96	80	100
92	75	96
96	84	98
92	82	100

(a) 这些商标的电池寿命有差异吗?

(b) 分析该实验的残差.

(c) 构造电池商标 2 的平均寿命的 95%区间估计. 构造电池商标 2 与商标 3 的平均寿命之差的 99%区间估计.

(d) 你愿意选用哪种商标的电池? 如果制造商要更换所有寿命低于 85 周的电池, 公司预期要更换的百分率有多大?

3.21 研究 4 种催化剂可能影响由 3 种成分构成的液体混合物中的一种成分的浓度问题. 利用完全随机化实验, 得到下述的浓度.

1	2	3	4
58.2	56.3	50.1	52.9
57.2	54.5	54.2	49.9
58.4	57.0	55.4	50.0
55.8	55.3		51.7
54.9			

(a) 4 种催化剂对浓度有相同的效应吗?

(b) 分析该实验的残差.

(c) 构造催化剂 1 的平均响应的 99%置信区间估计.

*3.22 进行一项实验, 研究 5 种绝缘材料的有效性. 加大电压水平以加速失效时间来检验每种材料的 4 个样本. 失效时间 (单位: 分钟) 如下表所示:

材料	失效时间 (分钟)			
1	110	157	194	178
2	1	2	4	18
3	880	1 256	5 276	4 355
4	495	7 040	5 307	10 050
5	7	5	29	2

(a) 5 种材料对失效时间均值有相同的效应吗?

(b) 作残差与预测响应的关系图. 构造残差的正态概率图. 这些图传达了什么信息?

(c) 基于 (b) 中的答案, 进行失效时间数据的另一种分析, 并得出合适的结论.

3.23 半导体厂研究了减少晶片粒子数的 3 种方法. 所有这 3 种方法都用 5 个不同的晶片检验并得到处理后的粒子数. 数据如下所示:

方法	粒 子 数				
1	31	10	21	4	1
2	62	40	24	30	35
3	53	27	120	97	68

(a) 所有这些方法对粒子数均值都有相同的效应吗?

(b) 作残差与预测响应的关系图. 构造残差的正态概率图. 这些和假设的有效性有关吗?

(c) 基于 (b) 中的答案, 进行粒子数的另一种分析, 并得出合适的结论.

3.24 考虑检验两个正态总体均值的相等性, 其方差未知但假定相等. 合适的检验法是合并 t 检验法. 证明合并 t 检验法等价于单因子方差分析.

3.25 证明线性组合 $\sum_{i=1}^{a} c_i y_i$ 的方差是 $\sigma^2 \sum_{i=1}^{a} n_i c_i^2$.

3.26 在一项固定效应实验中, 设 4 种处理的每一种有 n 个观测值. 设 Q_1^2, Q_2^2, Q_3^2 是正交对照的单自由度的分量. 证明 $SS_{处理} = Q_1^2 + Q_2^2 + Q_3^2$.

3.27 用 Bartlett 检验确定思考题 3.20 的等方差假定是否满足. 用 $\alpha = 0.05$. 通过检查残差图, 你能得到有关方差相等的同样结论吗?

3.28 用修正的 Levene 检验确定思考题 3.20 的等方差假定是否满足. 用 $\alpha = 0.05$. 通过检查残差图, 你能得到有关方差相等的同样结论吗?

3.29 再考虑思考题 3.16. 如果希望以概率至少为 0.90 检测平均响应时间为 10 微秒的最大差, 那么所需的样本量为多少? 讨论回答这个问题, 你如何求得 σ^2 的初步估计值.

*3.30 再考虑思考题 3.20.

(a) 如果希望以概率至少为 0.90 检测电池寿命为 10 小时的最大差异, 则所需的样本量为多少? 讨论回答这个问题, 你如何求得 σ^2 的初步估计值.

(b) 如果批次间的差异大得足以使一个观测值的标准差增加 25%, 并且我们希望以概率至少为 0.90 检测这一点, 则需要多大的样本量?

3.31 考虑思考题 3.20 中的实验. 如果要对两种电池的平均寿命的差构造一个 95%置信区间, 使其精确度为 ±2 周, 则对每种电池需要检测多少节?

*3.32 设 4 个正态总体的均值分别为 $\mu_1 = 50, \mu_2 = 60, \mu_3 = 50, \mu_4 = 60$. 从每个总体中应该取多少个观测值, 才能使得至少以 0.90 的概率拒绝均值相等的零假设? 假定 $\alpha = 0.05$, 误差方差的一个合理的估计量是 $\sigma^2 = 25$.

3.33 参考思考题 3.32.

(a) 如果实验误差方差的一个合理估计量是 $\sigma^2 = 36$, 你的答案将如何变化?

(b) 如果实验误差方差的一个合理估计量是 $\sigma^2 = 49$, 你的答案将如何变化?

(c) 这时, 关于 σ 的估计怎样影响样本量的精确度, 你会得出什么样的结论?

(d) 关于实践中如何使用一般的方法选择 n, 你有何建议?

3.34 参考 3.8 节描述的铝冶炼实验. 证明比值控制法不影响电解槽的平均电压. 画出残差的正态概率图. 画出残差与预测值的关系图. 图中显示出违反所作的基本假定吗?

*3.35 参考 3.8 节描述的铝冶炼实验. 对于电位噪音, 验证表 3.13 的方差分析. 检测通常的残差图并评价实验的有效性.

3.36 在一台数控机床上做一项实验来研究 4 种不同的进料速度, 这种机床是给飞机的辅助动力设备生产零部件的. 此实验的主管制造工程师知道, 所感兴趣的一个关键零件的尺寸受进料速度的影响. 但是, 经验表明, 只有分散效应很可能会出现. 也就是说, 改变进料速度不影响平均尺寸, 但可能会影响尺寸的变异性. 这位工程师对每种进料速度做了 5 次生产实验并求得关键尺寸 (以 10^{-3}mm 为单位) 的标准差, 数据如下. 假定所有的实验都按随机次序进行.

进料速度 (in/min)	生产实验				
	1	2	3	4	5
10	0.09	0.10	0.13	0.08	0.07
12	0.06	0.09	0.12	0.07	0.12
14	0.11	0.08	0.08	0.05	0.06
16	0.19	0.13	0.15	0.20	0.11

(a) 进料速度对关键尺寸的标准差有影响吗?

(b) 用此实验的残差来研究模型的适合性. 实验的有效性有问题吗?

3.37 考虑思考题 3.16 所示的数据.

(a) 写出这一问题的最小二乘正规方程组, 并用普通约束 $\left(\sum_{i=1}^{3} \hat{\tau}_i = 0\right)$ 解出 $\hat{\mu}$ 和 $\hat{\tau}_i$. 估计 $\tau_1 - \tau_2$.

(b) 用约束 $\hat{\tau}_3 = 0$ 来解 (a) 中的方程组. 估计量 $\hat{\mu}$ 与 $\hat{\tau}_i$ 和 (a) 中所得到的相同吗? 为什么? 估计 $\tau_1 - \tau_2$ 并与 (a) 中所得的进行比较. 关于在估计 τ_i 变量上的矛盾, 你能说出什么吗?

(c) 用正规方程组的这两个解来估计 $\mu + \tau_1$, $2\tau_1 - \tau_2 - \tau_3$ 以及 $\mu + \tau_1 + \tau_2$, 并比较在每种情形下所得的结果.

3.38 对例 3.1 中的实验应用一般回归显著性检验法. 说明此方法与通常的方差分析产生相同的结果.

*3.39 对思考题 3.7 的实验用 Kruskal-Wallis 检验法. 将所得的结论与从通常的方差分析法中所得的结论进行比较.

3.40 对思考题 3.8 中的实验用 Kruskal-Wallis 检验法. 所得的结果能与用通常的方差分析法所得的结果进行比较吗?

3.41 考虑例 3.1 中的实验. 设蚀刻率的最大观测值错记为 250Å/min. 这对通常的方差分析产生什么效应? 它对 Kruskal-Wallis 检验法产生什么效应?

第 4 章　随机化区组, 拉丁方, 以及有关的设计

本章纲要

4.1 随机化完全区组设计
4.1.1 RCBD 的统计分析
4.1.2 模型适合性检验
4.1.3 随机化完全区组设计的一些其他方面
4.1.4 估计模型参数和一般回归显著性检验
4.2 拉丁方设计
4.3 正交拉丁方设计
4.4 平衡不完全区组设计
4.4.1 BIBD 的统计分析
4.4.2 参数的最小二乘估计
4.4.3 BIBD 中内部信息的恢复
第 4 章补充材料
S4.1 RCBD 的相对效率
S4.2 部分平衡不完全区组设计
S4.3 Youden 方
S4.4 格点设计

4.1 随机化完全区组设计

在很多实验中, 由讨厌因子产生的变异性可能会影响结果. 一般地, 我们定义**讨厌因子** (nuisance factor) 为一个可能对响应产生影响的设计因子. 但是, 我们对这个效应并不感兴趣. 有时, 讨厌因子是**未知的**且**不可控制的**. 也就是说, 我们并不知道因子的存在, 但在进行实验时, 它甚至可能正在改变着因子水平. **随机化**是一种常用于抵制这种 "潜伏" 的讨厌因子的设计技术. 在其他情形中, 讨厌因子虽然是**已知**的但却是**不可控的**. 如果我们至少可以观测到讨厌因子影响了实验的每一轮中的值, 那么就可以用**协方差分析**(将在第 14 章中讨论) 作为它在统计分析中的补充. 当讨厌来源的变异性是**已知的**且**可控的**, 我们可以用所谓的**区组化**设计技术来系统地消除它在处理间统计比较中的效应. 区组化是一种十分重要的设计技术, 广泛应用于工业实验, 它也是本章的主要内容.

为了说明这种基本思想, 再考虑前面 2.5.1 节中讨论过的硬度检验试验. 假定我们想确定 4 只不同的压头在一台硬度试验机上是否得出不同的读数. 像这样的实验可能是测量仪能力研究的一部分. 机器将压头压入一块金属试件中去, 由压入的深度可确定试件的硬度. 实验者决定, 用洛氏硬度试验法对每只压头取 4 个观测值. 因为仅有一个因子 —— 压头类型, 所以, 一个完全随机化的单因子设计就由 $4 \times 4 = 16$ 次试验组成, 这 16 次试验中的每一次试验都被随机安排到一个*实验单元*(即, 一块金属试件), 然后观测所得的硬度读数. 这样一来, 在这个实验中就需要 16 块不同的金属试件, 每块作一次试验.

在这种设计情况下, 完全随机化实验存在一个**潜在**的严重问题. 如果金属试件取自不同炉次的铸锭, 可能造成它们的硬度略有不同, 那么, 实验单元 (试件) 就会对硬度数据观测值的变异性产生影响. 于是, 实验误差就不仅反映出随机误差, 还反映出试件间的变异性.

我们当然希望实验误差尽可能的小. 也就是说, 我们希望从实验误差中消除试件间的变异性. 为了做到这一点, 实验者在 4 块试件的每一块上对每只压头检验一次. 这种设计 (如表 4.1 所示) 叫做**随机化完全区组设计**(RCBD). "完全" 一词指的是每一区组 (试件) 包含了所有的处

理 (压头). 使用这种设计方法, 区组或试件形成了一个较齐性的实验单元, 在这些实验单元上比较压头的硬度. 从效果上说, 这种设计方案因为消除了试件间的变异性从而提高了压头间进行比较的精度. 在一个区组内, 4 只压头被检验的顺序是随机确定的. 这一设计问题和 2.5.1 节中曾讨论过的配对 t 检验法所提的问题有一定的相似性. 随机化完全区组设计是那个概念的推广.

表 4.1 硬度检验实验的随机化完全区组设计

试件 (区组)			
1	2	3	4
压头 3	压头 3	压头 2	压头 1
压头 1	压头 4	压头 1	压头 4
压头 4	压头 2	压头 3	压头 2
压头 2	压头 1	压头 4	压头 3

RCBD 是得到最广泛应用的实验设计之一. 适用 RCBD 的情况很多. 试验仪器或机器的单元间在其运行特性上常常也会有所不同, 这是一类典型的区组化因子. 原材料的批次、人以及时间, 也都是实验变异性的常见干扰源, 它们也是可以通过区组化而被系统控制的.

区组化对于不一定包含讨厌因子的情形也是有用的. 例如, 假定化学工程师想要研究催化剂进料率对聚合体黏性的影响. 她知道有许多因子 (例如原材料来源、温度、操作者以及原材料的纯度) 在整个过程中非常难以控制. 因此她决定用区组化的方法检验催化剂进料率因子, 这些不可控因子的某种组合构成一个区组. 事实上, 她用这些区组检验过程变量 (进料率) 相应于不易控制的条件的**稳健性** (robustness). 关于这一点的讨论, 详见 Coleman and Montgomery(1993).

4.1.1 RCBD 的统计分析

一般地, 假设我们有 a 个待比较的处理和 b 个区组. 随机化完全区组设计如图 4.1 所示. 在每个区组内, 每个处理都有一个观测值, 且进行试验时各种处理的次序是随机确定的. 因为只有在区组内, 处理才有随机性, 所以我们常说区组表示了关于**随机性的一种约束**.

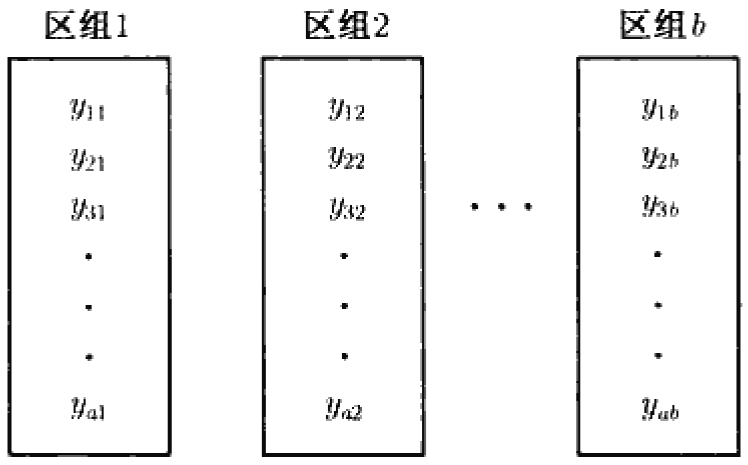


图 4.1 随机化完全区组设计

可以用几种方式写出关于 RCBD 的**统计模型**. 传统的模型是**效应模型** (effect model):

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b \end{cases} \tag{4.1}$$

其中 μ 是总均值, τ_i 是第 i 种处理的效应, β_j 是第 j 个区组的效应, ε_{ij} 是通常的 NID(0, σ^2) 随机误差项. 初步考虑处理与区组为固定因子. 正如第 3 章的单因子方差分析模型一样, RCBD 的效应模型是一个规定过多的模型. 我们通常将处理和区组效应定义为对总均值的偏差, 所以

$$\sum_{i=1}^{a}\tau_i = 0, \qquad \sum_{j=1}^{b}\beta_j = 0$$

也可以使用 RCBD 的**均值模型**, 即

$$y_{ij} = \mu_{ij} + \epsilon_{ij}\begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b \end{cases}$$

这里 $\mu_{ij} = \mu + \tau_i + \beta_j$. 但是在本章中, 我们将用效应模型 (4.1) 式.

在涉及 RCBD 的实验中, 我们的兴趣在于检验处理均值是否相等. 于是, 所感兴趣的假设是

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_a$$
$$H_1: 至少有一对\ \mu_i \neq \mu_j$$

因为第 i 个处理均值 $\mu_i = (1/b)\sum_{j=1}^{b}(\mu + \tau_i + \beta_j) = \mu + \tau_i$, 写出上述假设的一个等价方式是根据处理效应, 即

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1: 至少有一个\ \tau_i \neq 0$$

方差分析很容易扩展到 RCBD. 以 $y_{i.}$ 表示处理 i 下的所有观测值的总和, $y_{.j}$ 表示在区组 j 内所有观测值的总和, $y_{..}$ 表示全体观测值的总和, N=ab 是观测值的总个数. 用数学式表示为

$$y_{i.} = \sum_{j=1}^{b} y_{ij} \quad i = 1, 2, \cdots, a \tag{4.2}$$

$$y_{.j} = \sum_{i=1}^{a} y_{ij} \quad j = 1, 2, \cdots, b \tag{4.3}$$

$$y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij} = \sum_{i=1}^{a} y_{i.} = \sum_{j=1}^{b} y_{.j} \tag{4.4}$$

类似地, 以 $\overline{y}_{i.}$ 表示处理 i 下观测值的平均值, $\overline{y}_{.j}$ 表示区组 j 内观测值的平均值, $\overline{y}_{..}$ 表示全体观测值的总平均值, 即

$$\overline{y}_{i.} = y_{i.}/b, \quad \overline{y}_{.j} = y_{.j}/a, \quad \overline{y}_{..} = y_{..}/N \tag{4.5}$$

总校正平方和可表示为

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \overline{y}_{..})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}[(\overline{y}_{i.} - \overline{y}_{..}) + (\overline{y}_{.j} - \overline{y}_{..}) + (y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})]^2 \tag{4.6}$$

将 (4.6) 式右端项展开, 得

$$\begin{aligned}\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \overline{y}_{..})^2 &= b\sum_{i=1}^{a}(\overline{y}_{i.} - \overline{y}_{..})^2 + a\sum_{j=1}^{b}(\overline{y}_{.j} - \overline{y}_{..})^2 \\ &+ \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})^2 + 2\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{y}_{i.} - \overline{y}_{..})(\overline{y}_{.j} - \overline{y}_{..})\end{aligned}$$

$$+2\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{y}_{.j}-\overline{y}_{..})(y_{ij}-\overline{y}_{i.}-\overline{y}_{.j}+\overline{y}_{..})$$
$$+2\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{y}_{i.}-\overline{y}_{..})(y_{ij}-\overline{y}_{i.}-\overline{y}_{.j}+\overline{y}_{..})$$

由简单的但冗长的代数计算可证明三项交叉乘积都为零, 因此,

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\overline{y}_{..})^2=b\sum_{i=1}^{a}(\overline{y}_{i.}-\overline{y}_{..})^2+a\sum_{j=1}^{b}(\overline{y}_{.j}-\overline{y}_{..})^2+\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\overline{y}_{.j}-\overline{y}_{i.}+\overline{y}_{..})^2 \tag{4.7}$$

它表示总平方和的一个分解式. 这是 RCBD 的基本的方差分析等式用符号表示 (4.7) 式的平方和则是

$$SS_T=SS_{\text{处理}}+SS_{\text{区组}}+SS_E \tag{4.8}$$

因为有 N 个观测值, 所以 SS_T 有 $N-1$ 个自由度. 又因为有 a 种处理和 b 个区组, 所以 $SS_{\text{处理}}$ 和 $SS_{\text{区组}}$ 分别有 $a-1$ 和 $b-1$ 个自由度. 误差平方和恰是各单元间的平方和减去处理平方和与区组平方和, 其间有 ab 个单元, 具有 $ab-1$ 个自由度, 所以 SS_E 有 $ab-1-(a-1)-(b-1)=(a-1)(b-1)$ 个自由度. 还有, (4.8) 式右端项的自由度加起来等于左端项的. 因此, 用通常的关于误差的正态性假定, 利用定理 3.1 可证明 $SS_{\text{处理}}/\sigma^2$、$SS_{\text{区组}}/\sigma^2$ 以及 SS_E/σ^2 是独立分布的卡方随机变量. 每一平方和除以它的自由度就是均方. 如果处理和区组都是固定的, 均方的期望值可以证明是

$$E(MS_{\text{处理}})=\sigma^2+\frac{b\sum_{i=1}^{a}\tau_i^2}{a-1},\qquad E(MS_{\text{区组}})=\sigma^2+\frac{a\sum_{j=1}^{b}\beta_j^2}{b-1},\qquad E(MS_E)=\sigma^2$$

因此, 为了检验处理均值的等式, 可以用检验统计量

$$F_0=\frac{MS_{\text{处理}}}{MS_E}$$

当零假设为真时, 它的分布是 $F_{a-1,(a-1)(b-1)}$. 拒绝域是 F 分布的上尾部, 且当 $F_0>F_{a,a-1,(a-1)(b-1)}$ 时, 拒绝 H_0.

我们也有兴趣于比较区组均值, 因为, 如果这些均值没有大的差异, 则在将来的实验中, 区组化就没必要了. 从期望均方来说, 将统计量 $F_0=MS_{\text{区组}}/MS_E$ 与 $F_{a,a-1,(a-1)(b-1)}$ 比较, 似乎就可以检验假设 $H_0:\beta_j=0$. 然而, 随机化仅应用在区组内部的处理上. 也就是说, 区组表示一个**随机化约束**. 那么, 这一点在统计量 $F_0=MS_{\text{区组}}/MS_E$ 上有什么影响呢? 在处理这个问题上是存在一些差别的. 例如, Box, Hunter, and Hunter(1978) 指出, 通常的方差分析 F 检验法, 只需基于随机化就可说明是合理的①, 而不再需要正态性假定. 他们进一步观测到, 因为随机化约束, 比较区组均值的检验法不再显现出这种合理性, 但是, 如果误差是 NID(0, σ^2) 时, 则统计量 $F_0=MS_{\text{区组}}/MS_E$ 可以用来比较区组均值. 另一方面, Anderson and McLean(1974) 断言, 随机化约束使这一统计量在比较区组均值中不是有意义的检验法, 这个 F 比值实际上

① 确切说, 标准理论的 F 分布是由计算 F_0 而得到的随机化分布的一种近似, 这里的 F_0 的计算涉及处理响应的每一种可能的赋值.

是检验区组均值的等式加随机化约束 [他们称之为约束误差，更详细的论述参阅 Anderson and McLean(1974)].

那么，在实践中，我们怎样做呢？因为正态性假定通常是有问题的，$F_0 = MS_{区组}/ MS_E$ 作为区组均值等式的精确的 F 检验法就不是一种好的一般实践. 因此，我们从方差分析表中排除这一 F 检验法. 然而，作为研究区组变量效应的一种近似方法，检验 $MS_{区组}$ 对 MS_E 的比值当然是合理的. 当这一比值较大时，表明区组因子有大的影响，因而，由区组化所得的噪音简化可能有助于改善处理均值比较的精度.

这一方法通常用如表 4.2 所示的方差分析表概括出来. 通常可以用一个统计软件包进行计算. 然而，计算平方和的公式可以通过分解 (4.7) 式中的等式

$$y_{ij} - \overline{y}_{..} = (\overline{y}_{i.} - \overline{y}_{..}) + (\overline{y}_{.j} - \overline{y}_{..}) + (y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})$$

表 4.2　随机化完全区组设计的方差分析

方差来源	平方和	自由度	均方	F_0
处理	$SS_{处理}$	$a-1$	$\frac{SS_{处理}}{a-1}$	$\frac{MS_{处理}}{MS_E}$
区组	$SS_{区组}$	$b-1$	$\frac{SS_{区组}}{b-1}$	
误差	SS_E	$(a-1)(b-1)$	$\frac{SS_E}{(a-1)(b-1)}$	
总和	SS_T	$N-1$		

得到. 这些量可以在工作表 (如 Excel) 的列中进行计算. 对每列进行平方再求和得到平方和. 另外，计算公式可以用处理和与区组和来表示. 这些计算公式是

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{4.9}$$

$$SS_{处理} = \frac{1}{b}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N} \tag{4.10}$$

$$SS_{区组} = \frac{1}{a}\sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{N} \tag{4.11}$$

而误差平方和由减法得出为

$$SS_E = SS_T - SS_{处理} - SS_{区组} \tag{4.12}$$

例 4.1　某医疗器械工厂生产供移植用的人造血管. 通过将聚四氟乙烯 (PTFE) 树脂块借助于润滑剂挤压成管子制造人造血管. 生产流程中一些管子的外表面经常会出现小的、硬的突出物. 这些缺陷通常称为“疵点”. 这些缺陷导致产品不合格.

人造血管的产品开发者怀疑挤压压强导致了疵点的产生，从而想到做一个实验来验证这个假设. 然而，生产用的树脂由外来供应商按批次送到医疗器械工厂. 工程师也怀疑各批次之间有显著性差异，因为尽管材料的有关参数 (比如，分子量、平均微粒尺寸、保持力、峰高比等参数) 应该是一致的，但材料可能并不因为树脂供应商的生产过程变化以及材料的自然变化而变化. 因此，产品开发者决定考虑将树脂的批次作为区组，运用完全随机化区组设计，研究 4 个不同水平的挤压压强对疵点的效应. 该 RCBD 如表 4.3 所示. 有 4 个水平的挤压压强 (处理)，6 个批次的树脂 (区组). 注意到在每个区组内的挤压压强的试验次序是随机的. 响应变量是产率，即没有疵点的成品的百分比.

表 4.3 血管移植实验的随机化完全区组设计

挤压压强 (PSI)	树脂批次 (区组) 1	2	3	4	5	6	处理总和
8 500	90.3	89.2	98.2	93.9	87.4	97.9	556.9
8 700	92.5	89.5	90.6	94.7	87.0	95.8	550.1
8 900	85.5	90.8	89.6	86.2	88.0	93.4	533.5
9 100	82.5	89.5	85.6	87.4	78.9	90.7	514.6
区组总和	350.8	359.0	364.0	362.2	341.3	377.8	$y_{..} = 2\ 155.1$

为进行方差分析, 先计算下面的平方和:

$$SS_T = \sum_{i=1}^{4}\sum_{j=1}^{6} y_{ij}^2 - \frac{y_{..}^2}{N} = 193\ 999.31 - \frac{(2\ 155.1)^2}{24} = 480.31$$

$$SS_{处理} = \frac{1}{b}\sum_{i=1}^{4} y_{i.}^2 - \frac{y_{..}^2}{N} = \frac{1}{6}[(556.9)^2 + (550.1)^2 + (533.5)^2 + (514.6)^2] - \frac{(2\ 155.1)^2}{24} = 178.17$$

$$SS_{区组} = \frac{1}{a}\sum_{j=1}^{6} y_{.j}^2 - \frac{y_{..}^2}{N} = \frac{1}{4}[(350.8)^2 + (359.0)^2 + \cdots + (377.8)^2] - \frac{(2\ 155.1)^2}{24} = 192.25$$

$$SS_E = SS_T - SS_{处理} - SS_{区组} = 480.31 - 178.17 - 192.25 = 109.89$$

方差分析见表 4.4. 用 $\alpha = 0.05$, F 的临界值为 $F_{0.05,3,15} = 3.29$. 因为 $8.11 > 3.29$, 我们得出结论: 挤压压强影响了产品的均值. 检验的 P 值也很小. 同样, 树脂批次 (区组) 似乎也有显著性差异, 因为区组的均方相对于误差也很大.

表 4.4 人造血管移植实验的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
处理 (挤压压强)	178.17	3	59.39	8.11	0.001 9
区组 (批次)	192.25	5	38.45		
误差	109.89	15	7.33		
总和	480.31	23			

如果我们不知道随机化区组设计, 则观测那些可能得到的结果是很有趣的. 假设这个实验按照完全随机设计来做, 并且 (碰巧) 同样的实验结果如表 4.3 所示. 这些数据作为完全随机单因子设计的不正确的分析如表 4.5 所示:

表 4.5 完全随机设计的血管移植实验的不正确分析

方差来源	平方和	自由度	均方	F_0	P 值
挤压压强	178.17	3	59.39	3.95	0.023 5
误差	302.14	20	15.11		
总和	480.31	23			

因为 P 值小于 0.05, 我们仍拒绝原假设, 得到挤压压强显著影响平均产率的结论. 但是, 注意到均方误差成倍增加, 从 RCBD 的 7.33 增加到 15.11. 所有的变化是因为将区组放在误差项里. 容易理解为什么我们有时称 RCBD 为降噪设计技术, 因为它有效地增加了数据中的信噪比, 或者说, 它提高了处理均值比较的精度. 这个例子也说明了这一点. 如果实验者在必须区组化时, 没有区组化, 那么误差效应被夸大了, 很有可能夸大太多, 导致处理均值间的重要差异不能被识别出来.

1. 计算机输出示例

对于例 4.1 的人造血管实验, 用 Design-Expert 得出的计算机输出浓缩后如图 4.2 所示. 每一个处理的样本均值都在输出中显示. 在 8 500 psi, 平均产率为 $\bar{y}_{1.} = 92.82$; 在 8 700 psi,

平均产率为 $\overline{y}_{2.} = 91.68$; 在 8 900 psi, 平均产率为 $\overline{y}_{3.} = 88.92$; 在 9 100 psi, 平均产率为 $\overline{y}_{4.} = 85.77$. 用样本均值估计处理均值 $\mu_1, \mu_2, \mu_3, \mu_4$. 模型的残差显示在计算机输出的底部. 残差由下式计算

$$e_{ij} = y_{ij} - \widehat{y}_{ij}$$

Response: Yield
ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Block	192.25	5	38.45		
Model	178.17	3	59.39	8.11	0.0019
A	*178.17*	*3*	*59.39*	*8.11*	*0.0019*
Residual	109.89	15	7.33		
Cor Total	480.31	23			

Std. Dev.	2.71	R-Squared	0.6185
Mean	89.80	Adj R-Squared	0.5422
C.V.	3.01	Pred R-Squared	0.0234
PRESS	281.31	Adeq Precision	9.759

Treatment Means (Adjusted, If Necessary)

	Estimated Mean	Standard Error
1-8500	92.82	1.10
2-8700	91.68	1.10
3-8900	88.92	1.10
4-9100	85.77	1.10

Treatment	Mean Difference	DF	Standard Error	t for H_0 Coeff=0	Prob > |t|
1 vs 2	1.13	1	1.56	0.73	0.4795
1 vs 3	3.90	1	1.56	2.50	0.0247
1 vs 4	7.05	1	1.56	4.51	0.0004
2 vs 3	2.77	1	1.56	1.77	0.0970
2 vs 4	5.92	1	1.56	3.79	0.0018
3 vs 4	3.15	1	1.56	2.02	0.0621

Diagnostics Case Statistics

Standard Order	Actual Value	Predicted Value	Residual	Leverage	Student Residual	Cook's Distance	Outlier t	Run Order
1	90.30	90.72	−0.42	0.375	−0.197	0.003	−0.190	1
2	89.20	92.77	−3.57	0.375	−1.669	0.186	−1.787	6
3	98.20	94.02	4.18	0.375	1.953	0.254	2.185	9
4	93.90	93.57	0.33	0.375	0.154	0.002	0.149	13
5	87.40	88.35	−0.95	0.375	−0.442	0.013	−0.430	19
6	97.90	97.47	0.43	0.375	0.201	0.003	0.194	23
7	92.50	89.59	2.91	0.375	1.361	0.124	1.405	4
8	89.50	91.64	−2.14	0.375	−0.999	0.067	−0.999	5
9	90.60	92.89	−2.29	0.375	−1.069	0.076	−1.075	10
10	94.70	92.44	2.26	0.375	1.057	0.075	1.062	16
11	87.00	87.21	−0.21	0.375	−0.099	0.001	−0.096	20
12	95.80	96.34	−0.54	0.375	−0.251	0.004	−0.243	21
13	85.50	86.82	−1.32	0.375	−0.617	0.025	−0.604	3
14	90.80	88.87	1.93	0.375	0.902	0.054	0.896	8
15	89.60	90.12	−0.52	0.375	−0.243	0.004	−0.236	12
16	86.20	89.67	−3.47	0.375	−1.622	0.175	−1.726	15
17	88.00	84.45	3.55	0.375	1.661	0.184	1.776	17
18	93.40	93.57	−0.17	0.375	−0.080	0.000	−0.077	22
19	82.50	83.67	−1.17	0.375	−0.547	0.020	−0.534	2
20	89.50	85.72	3.78	0.375	1.766	0.208	1.917	7
21	85.60	86.97	−1.37	0.375	−0.641	0.027	−0.628	11
22	87.40	86.52	0.88	0.375	0.411	0.011	0.399	14
23	78.90	81.30	−2.40	0.375	−1.120	0.084	−1.130	18
24	90.70	90.42	0.28	0.375	0.130	0.001	0.126	24

Note: Predicted values include block corrections.

图 4.2　例 4.1 的 Design-Expert 输出 (已浓缩)

如我们将在后面介绍的, 拟合值 $\hat{y}_{ij} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$. 所以,

$$e_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..} \tag{4.13}$$

在 4.1.2 节中, 我们将说明残差是如何用在**模型适合性检验**中的.

2. 多重比较

如果 RCBD 中的处理是固定的, 并且分析表明处理均值有显著性差异, 则实验者通常会对多重比较感兴趣, 以便发现哪些处理的均值有差异. 第 3 章 (3.5 节) 所讨论的任一种多重比较方法都可用来达到这一目标. 只要在 3.5 节的公式中简单地用区组数 (b) 代替单因子完全随机化设计的重复次数 (n) 就可以了. 还要记住要用随机化区组设计中误差的自由度数 $[(a-1)(b-1)]$ 来代替完全随机化设计的误差自由度 $[a(n-1)]$.

图 4.2 中的 Design-Expert 输出说明了 Fisher LSD 方法, 我们得到的结论是 $\mu_1 = \mu_2$, 因为 P 值非常大. 此外, μ_1 和其他所有均值不同. $H_0: \mu_2 = \mu_3$ 的 P 值是 0.097, 所以有根据表明 $\mu_2 \neq \mu_3$. 由 P 值等于 0.001 8 , 得 $\mu_3 \neq \mu_4$. 总之, 我们得出结论: 较小的挤压压强导致较少的缺陷.

也可以用第 3 章 (3.5.1 节) 的图示法来比较 4 种挤压压强的平均产率. 图 4.3 标出了例 4.1 的 4 个均值点, 此图是按 t 分布乘上一个尺度因子 $\sqrt{MS_E/b} = \sqrt{7.33/6} = 1.10$ 来绘制的. 图中表明, 两个最低的压强产生了相同的平均产率, 但是, 对于 8 700psi 和 8 900psi(μ_2 和 μ_3) 的平均产率也类似. 最高的压强 (9 100psi) 得到的平均产率显然不同于其他所有的均值. 该图有助于解释实验的结果以及图 4.2 中 Design-Expert 输出的 Fisher LSD 计算.

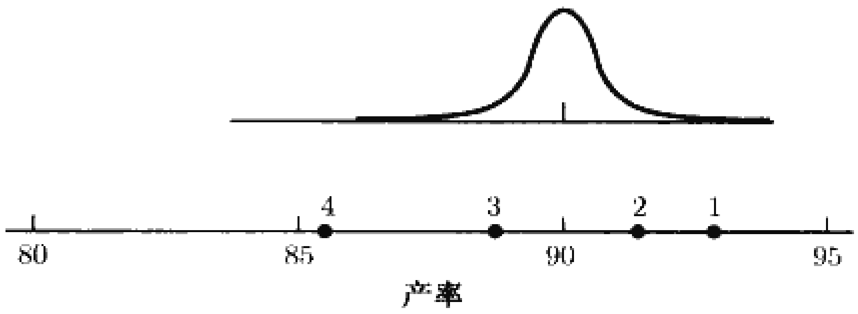


图 4.3 4 种挤压压强的平均产率, 按 t 分布乘上一个尺度因子 $\sqrt{MS_E/b} = \sqrt{7.33/6} = 1.10$

4.1.2 模型适合性检验

前面讨论过检验假定模型适合性的重要性. 一般说来, 我们应该警惕那些由正态性假定、处理或区组的不等的误差方差以及区组–处理交互作用引起的潜在问题. 如同在完全随机化设计中那样, 残差分析是用于这类诊断性检验的主要工具. 例 4.1 的随机化区组设计的残差被列在图 4.2 中的 Design-Expert 输出的底部.

这些残差的正态概率图如图 4.4 所示. 既没有非正态性的严重标志, 也没有任何存在离群值的证据. 图 4.5 画出了拟合值 $\hat{y}_{ij}$ 与残差的关系图. 残差的大小与拟合值 $\hat{y}_{ij}$ 之间没有什么关系. 这张图没有显示出令人特别感兴趣的地方. 图 4.6 描绘的是处理 (挤压压强) 与残差的关系图以及树脂的批次 (即区组) 与残差的关系图. 这些图可能非常有益. 如果一特定处理的残差较为分散, 则表示这一处理与其余的处理相比, 比较容易得出不稳定的响应. 较为分散的残差标志

着区组不齐性. 不过, 在我们的例子中, 图 4.6 并未表明处理方差不齐性, 但却表明 6 个批次的产量有较小的不同. 然而, 如果其他所有的残差图都令人满意, 我们就可以忽视这一点.

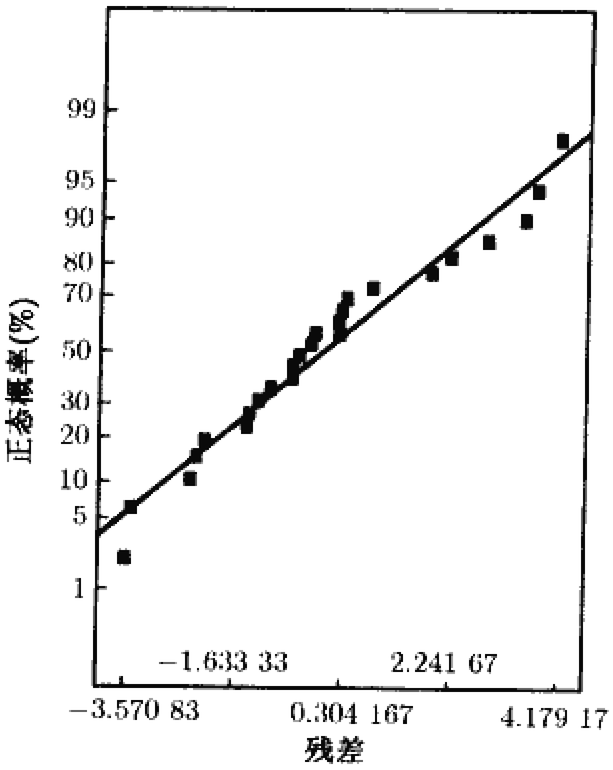


图 4.4　例 4.1 的残差的正态概率图

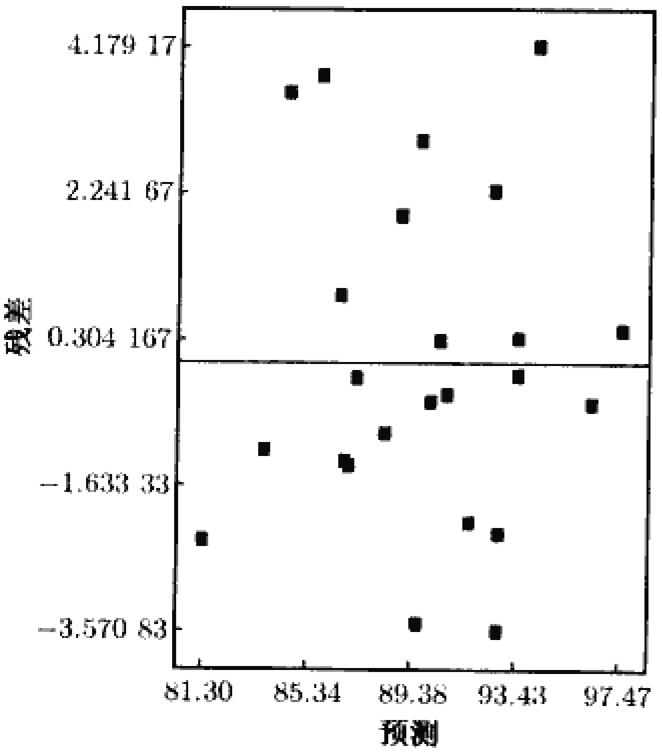


图 4.5　例 4.1 的残差与 y_{ij} 的关系图

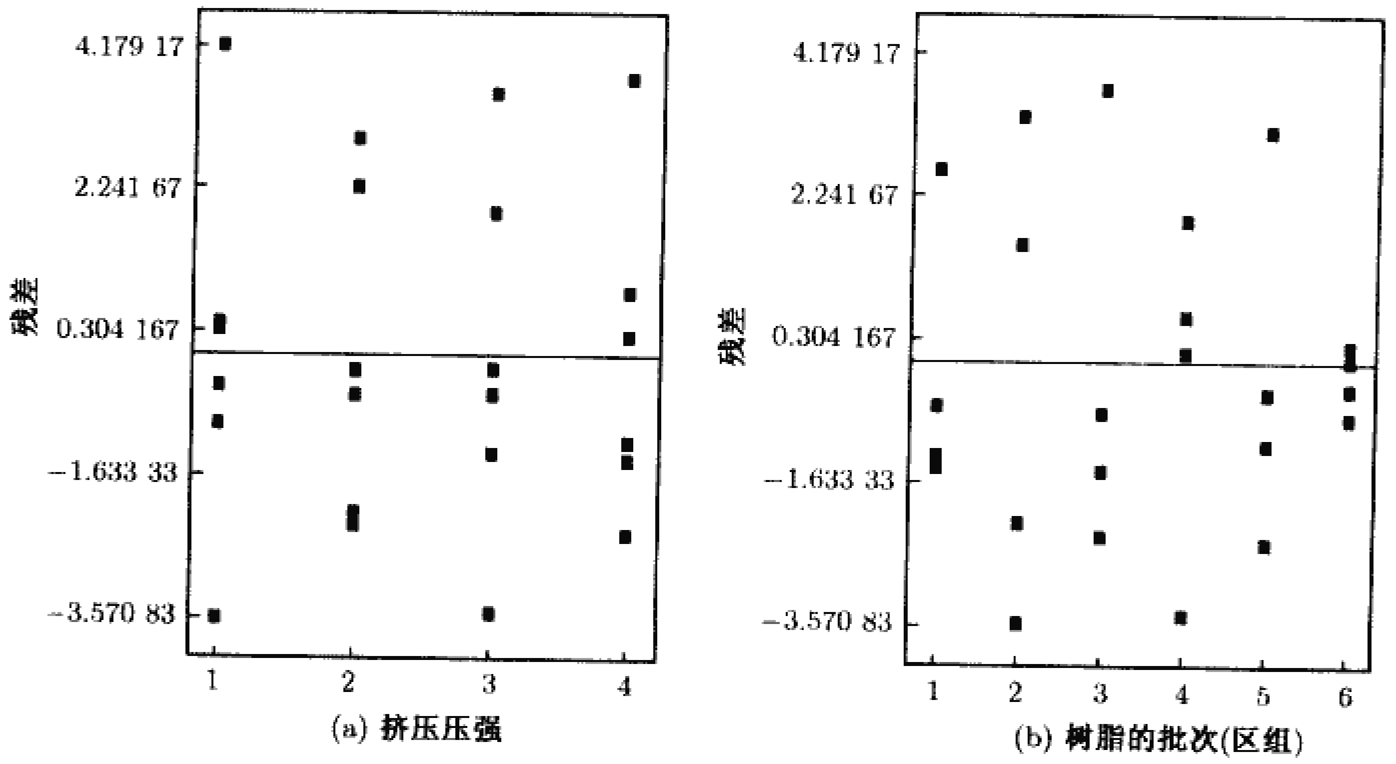


图 4.6　例 4.1 的残差与挤压压强 (处理) 以及树脂的批次 (区组) 的关系图

有时, 残差与 $\hat{y}_{ij}$ 的关系图是一条曲线. 例如, 对于低的 $\hat{y}_{ij}$ 值, 存在着产生负残差的趋向; 对中等的 $\hat{y}_{ij}$ 值, 存在着产生正残差的趋向; 对高的 $\hat{y}_{ij}$ 值, 存在着产生负残差的趋向. 这类图像暗示着区组与处理间存在着**交互作用**. 如果这类图像出现, 就要利用变换来消除或最小化交互作用. 第 5 章 (5.3.7 节) 中给出了一个统计检验法, 可以用来检验随机化区组设计中可能存在的交互作用.

4.1.3 随机化完全区组设计的一些其他方面

1. 随机化区组模型的可加性

用于随机化区组设计的线性模型

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$$

是可加的. 这意味着, 例如, 当第 1 种处理引起期望响应增加 5 个单位 ($\tau_1 = 5$) 以及第 1 个区组增加期望响应 2 个单位 ($\beta_1 = 2$) 时, 则处理 1 与区组 1 一起增加的期望响应 $\mathrm{E}(y_{11}) = \mu + \tau_1 + \beta_1 = \mu + 5 + 2 = \mu + 7$. 一般来说, 处理 1 就在总均值与区组效应的和上通常增加期望响应 5 个单位.

尽管这一简单可加模型经常用到, 但也存在它不适用的情况. 例如, 假定我们用 6 批原材料来比较一种化学产品的 4 种配方, 原材料的批次被考虑作区组. 如果批次 2 的一种杂质对配方 2 有不利影响而导致不寻常的低产量, 但对其他配方没有影响, 则在配方 (或处理) 和批次 (或区组) 之间就存在**交互作用** (interaction). 类似地, 当用错误的尺度来度量响应时, 在处理和区组之间也会出现交互作用. 因此, 在原来单位下的倍数关系为

$$\mathrm{E}(y_{ij}) = \mu\tau_i\beta_j$$

但在对数尺度意义下, 该关系是线性的或可加的, 例如,

$$\ln \mathrm{E}(y_{ij}) = \ln \mu + \ln \tau_i + \ln \beta_j$$

即

$$\mathrm{E}(y_{ij}^*) = \mu^* + \tau_i^* + \beta_j^*$$

虽然这类交互作用可以通过某一变换来消除, 但并不是所有的交互作用都那么容易对付. 例如, 前述的配方–批次交互作用就不能用变换来消除. 残差分析和其他诊断检验法会有助于检验非可加性.

当出现交互作用时, 会严重影响方差分析, 甚至可能使方差分析无效. 一般来说, 交互作用的出现会提高误差均方值并且反过来会影响到处理均值的比较. 若因子及其可能的交互作用都需要研究, 则必须用**析因设计**(factorial design). 这些设计将在第 5 章至第 9 章中进一步讨论.

2. 随机处理与区组

虽然我们描述了将处理和区组考虑为固定因子时的检验方法, 但如果处理或区组 (或两者都是) 是随机的, 那么可用同样的分析方法. 当然, 在对结果进行解释时, 要作相应的更改. 例如, 如果区组是随机的, 则我们希望处理间的比较相对于**区组的总体**是相同的. 期望均方值亦有相应的更改. 例如, 如果区组是独立随机变量且有共同的方差, 则 $\mathrm{E}(MS_{区组}) = \sigma^2 + a\sigma_\beta^2$, 其中 σ_β^2 是区组效应的方差分量. 不过, 在任一事件中, $\mathrm{E}(MS_{处理})$ 总与任一区组效应无关, 并且处理间变异性的检验统计量总是 $F_0 = MS_{处理}/MS_E$.

在区组是随机的情况中, 如果出现处理 - 区组交互作用, 则关于处理均值的检验不受交互作用的影响. 这是因为, 处理和误差的期望均方值二者都含有交互作用效应, 因此, 检验处理均值

差就可以像通常比较处理均方值与误差均方值那样来做. 但这一方法得不到关于交互作用的任何信息.

3. 样本量的选择

在使用 RCBD 时，选取**样本量**，或者选择用于实验的**区组数**是一项重要的决策. 增加区组数就会增大重复的次数和误差的自由度数，会使设计更加灵敏. 第 3 章 (3.7 节) 关于完全随机化单因子实验讲解了有关选取实验的重复次数的方法问题，其中的任一种方法都可以直接用于随机化区组设计. 对于固定因子情形，可利用附录图 V 的抽检特性曲线

$$\Phi^2 = \frac{b\sum_{i=1}^{a}\tau_i^2}{a\sigma^2} \tag{4.14}$$

其中分子的自由度是 $a-1$，分母的自由度是 $(a-1)(b-1)$.

例 4.2　考虑例 4.1 描述的人造血管的 RCBD. 如果想要以相当高的概率来检验产率的最大真实差为 6，而误差的标准差估计是 $\sigma=3$，我们希望确认合适的区组数，使设计更灵敏. 由 (3.45) 式，Φ^2 的最小值是 (将 n 写为区组数 b)

$$\Phi^2 = \frac{bD^2}{2a\sigma^2}$$

其中 D 是我们想要检验的差的最大值. 于是

$$\Phi^2 = \frac{b(6)^2}{2(4)(3)^2} = 0.5b$$

如果用 b=5 个区组，则 $\Phi=\sqrt{0.5b}=\sqrt{0.5(5)}=1.58$，其误差自由度 $(a-1)(b-1)=3(4)=12$. 在附录的图 V 中用 $\mu_1=a-1=3$ 和 $\alpha=0.05$，找到这一设计的 β 风险近似为 0.55(势 $=1-\beta=0.45$). 如果用 b=6 个区组，则 $\Phi=\sqrt{0.5b}=\sqrt{0.5(6)}=1.73$，而误差自由度为 $(a-1)(b-1)=3(5)=15$，对应的近似的 β 风险是 0.4(势 $=1-\beta=0.6$). 由于购买多批树脂的花费大，且实验的费用又高，所以实验者决定用 6 个区组，尽管势大约只有 0.6(确实许多势只有 0.5 或更高的实验做得很好).

4. 缺失值的估计

在用 RCBD 时，有时，某一个区组的一个观测值缺失. 由于不小心或犯错误或失去控制，这种情况是会发生的. 例如，一个实验单元出现了不可避免的损伤等. 一个缺失的观测值给分析带来了一个新问题，因为处理不再与**区组正交**了，也就是说，并不是每一个处理都在每一个区组中出现. 对于缺失值的问题，有两个一般的近似方法. 一个**近似分析法**是，估计这一缺失观测值，并把它作为真实数据进行通常的方差分析，只是将误差自由度减小 1. 这一种近似分析法是本节的主题. 另一个是一个**精确的分析法**，将在 4.1.4 节中讨论.

设处理 i 在区组 j 中的观测值 y_{ij} 缺失. 用 x 表示这一缺失值. 作为一种解释，假定在例 4.1 的人造血管试验中，当挤压机器用 8 700 psi 操作第 4 批材料时出现问题，并且没有得到该点的数据. 就会出现如表 4.6 那样的数据.

一般地，以 y'_{ij} 表示 y_{ij} 缺失时其他数据的总和，$y'_{i.}$ 表示有一个缺失值的处理的总和，$y'_{.j}$ 表示有一个缺失值时区组的总和. 假设我们希望被估计的那个缺失值 x 对误差平方和有最小的贡献. 因为 $SS_E=\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2$，这个问题就等价于去选取 x，使得

表 4.6 含有一个缺失数据的血管移植实验的随机化完全区组设计

挤压压强 (psi)	树脂的批次 (区组)						处理总和
	1	2	3	4	5	6	
8 500	90.3	89.2	98.2	93.9	87.4	97.9	556.9
8 700	92.5	89.5	90.6	x	87.0	95.8	455.4
8 900	85.5	90.8	89.6	86.2	88.0	93.4	533.5
9 100	82.5	89.5	85.6	87.4	78.9	90.7	519.6
区组总和	350.8	359.0	364.0	267.5	341.3	377.8	$y'_{..}=2\ 060.4$

$$SS_E=\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij}^2-\frac{1}{b}\sum_{i=1}^{a}\left(\sum_{j=1}^{b}y_{ij}\right)^2-\frac{1}{a}\sum_{j=1}^{b}\left(\sum_{i=1}^{a}y_{ij}\right)^2+\frac{1}{ab}\left(\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij}\right)^2$$

即

$$SS_E=x^2-\frac{1}{b}(y'_{i.}+x)^2-\frac{1}{a}(y'_{.j}+x)^2+\frac{1}{ab}(y'_{..}+x)^2+R \tag{4.15}$$

最小化, 其中 R 包括所有与 x 无关的项. 由 $\mathrm{d}SS_E/\mathrm{d}x=0$ 得

$$x=\frac{ay'_{i.}+by'_{.j}-y'_{..}}{(a-1)(b-1)} \tag{4.16}$$

把它作为缺失值的估计值.

对于表 4.6 中的数据, 得 $y'_{2.}=455.4$, $y'_{.4}=267.5$, $y'_{..}=2\ 060.4$. 因此, 由 (4.16) 式

$$x\equiv y_{24}=\frac{4(455.4)+6(267.5)-2\ 060.4}{(3)(5)}=91.08$$

现在, 用 $y_{24}=91.08$ 来进行通常的方差分析, 并将误差自由度减少 1. 方差分析如表 4.7 所示, 请将这一近似分析法与由完全数据组 (表 4.4) 所得的结果比较一下.

表 4.7 含有一个缺失数据的例 4.1 的近似方差分析

方差来源	平方和	自由度	均方	F_0	P- 值
挤压压强	166.14	3	55.38	7.63	0.002 9
原材料批次	189.52	5	37.90		
误差	101.70	14	7.26		
总和	457.36	23			

如果有几个观测值缺失, 则可以把误差平方和写为这些缺失值的函数, 对每一缺失值求导数, 令其为零, 然后解这一方程组, 求得缺失值的估计值. 也可以由 (4.16) 式用迭代法来估计缺失值. 为了解释迭代法, 设有两个缺失值. 任意估计第一个缺失值, 然后用这一数据作为真实值并与 (4.16) 式一起去估计第二个. 现在, (4.16) 式再用来重新估计第一个缺失值, 此后, 又可以重新估计第二个. 这一过程继续进行下去, 直至收敛为止. 在任何缺失值问题中, 误差自由度对每一个缺失值都减少 1.

4.1.4 估计模型参数和一般回归显著性检验

如果处理和区组都是固定的, 则可以用最小二乘法来估计随机化完全区组设计模型的参数. 我们采用的线性统计模型是

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b \end{cases} \tag{4.17}$$

利用 3.9.2 节中对实验设计模型求正规方程组的规则, 得

$$\begin{aligned}
\mu &: ab\hat{\mu} + b\hat{\tau}_1 + b\hat{\tau}_2 + \cdots + b\hat{\tau}_a + a\hat{\beta}_1 + a\hat{\beta}_2 + \cdots + a\hat{\beta}_b = y_{..} \\
\tau_1 &: \ b\hat{\mu} + b\hat{\tau}_1 \qquad\qquad\qquad + \hat{\beta}_1 + \hat{\beta}_2 + \cdots + \hat{\beta}_b = y_{1.} \\
\tau_2 &: \ b\hat{\mu} \qquad + b\hat{\tau}_2 \qquad\qquad + \hat{\beta}_1 + \hat{\beta}_2 + \cdots + \hat{\beta}_b = y_{2.} \\
&\vdots \\
\tau_a &: \ b\hat{\mu} \qquad\qquad b\hat{\tau}_a + \hat{\beta}_1 + \hat{\beta}_2 + \cdots + \hat{\beta}_b = y_{a.} \\
\beta_1 &: \ a\hat{\mu} + \hat{\tau}_1 + \hat{\tau}_2 + \cdots + \hat{\tau}_a + a\hat{\beta}_1 \qquad\qquad = y_{.1} \\
\beta_2 &: \ a\hat{\mu} + \hat{\tau}_1 + \hat{\tau}_2 + \cdots + \hat{\tau}_a \qquad + a\hat{\beta}_2 \qquad = y_{.2} \\
&\vdots \\
\beta_b &: \ a\hat{\mu} + \hat{\tau}_1 + \hat{\tau}_2 + \cdots + \hat{\tau}_a \qquad\qquad + a\hat{\beta}_b = y_{.b}
\end{aligned} \tag{4.18}$$

注意到在 (4.18) 式中, 将第 2 个到第 $(a+1)$ 个方程加起来, 就是第一个方程. 而将其后的 b 个方程加起来, 也是第一个方程. 这样一来, 在这些正规方程组中, 有两个方程和其他方程线性相关. 这就是说, 要解方程 (4.18), 必须添加两个约束. 通常的约束是

$$\sum_{i=1}^{a} \hat{\tau}_i = 0, \qquad \sum_{j=1}^{b} \hat{\beta}_j = 0 \tag{4.19}$$

使用这些约束, 正规方程组就简单得多了. 事实上, 方程组将变为

$$\begin{aligned}
ab\hat{\mu} &= y_{..} \\
b\hat{\mu} + b\hat{\tau}_i &= y_{i.}, \qquad i = 1, 2, \cdots, a \\
a\hat{\mu} + a\hat{\beta}_j &= y_{.j}, \qquad j = 1, 2, \cdots, b
\end{aligned} \tag{4.20}$$

其解是

$$\begin{aligned}
\hat{\mu} &= \overline{y}_{..} \\
\hat{\tau}_i &= \overline{y}_{i.} - \overline{y}_{..}, \qquad i = 1, 2, \cdots, a \\
\hat{\beta}_j &= \overline{y}_{.j} - \overline{y}_{..}, \qquad j = 1, 2, \cdots, b
\end{aligned} \tag{4.21}$$

用 (4.21) 式的正规方程的解, 得出 y_{ij} 的估计即拟合值为

$$\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j = \overline{y}_{..} + (\overline{y}_{i.} - \overline{y}_{..}) + (\overline{y}_{.j} - \overline{y}_{..}) = \overline{y}_{i.} + \overline{y}_{.j} - \overline{y}_{..}$$

这一结果在前面计算随机化区组设计中的残差时就已经在 (4.13) 式中用过.

一般回归的显著性检验法可以用来进行随机化完全区组设计的方差分析. 用由 (4.21) 式给出的正规方程组的解, 拟合全模型的简化平方和是

$$\begin{aligned}
R(\mu, \tau, \beta) &= \hat{\mu} y_{..} + \sum_{i=1}^{a} \hat{\tau}_i y_{i.} + \sum_{j=1}^{b} \hat{\beta}_j y_{.j} = \overline{y}_{..} y_{..} + \sum_{i=1}^{a} (\overline{y}_{i.} - \overline{y}_{..}) y_{i.} + \sum_{j=1}^{b} (\overline{y}_{.j} - \overline{y}_{..}) y_{.j} \\
&= \frac{y_{..}^2}{ab} + \sum_{i=1}^{a} \overline{y}_{i.} y_{i.} - \frac{y_{..}^2}{ab} + \sum_{j=1}^{b} \overline{y}_{.j} y_{.j} - \frac{y_{..}^2}{ab} = \sum_{i=1}^{a} \frac{y_{i.}^2}{b} + \sum_{j=1}^{b} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}
\end{aligned}$$

它有 $a+b-1$ 个自由度, 且误差平方和是

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - R(\mu,\tau,\beta) = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \sum_{i=1}^{a}\frac{y_{i.}^2}{b} - \sum_{j=1}^{b}\frac{y_{.j}^2}{a} + \frac{y_{..}^2}{ab}$$
$$= \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})^2$$

它有 $(a-1)(b-1)$ 个自由度. 请将这最后一个等式与 (4.7) 式中的 SS_E 进行比较.

检验假设 $H_0: \tau_i = 0$, 相应的**简化模型**是

$$y_{ij} = \mu + \beta_j + \epsilon_{ij}$$

它恰巧是单因子方差分析. 与 (3.5) 式类似, 拟合简化模型的简化平方和是

$$R(\mu,\beta) = \sum_{j=1}^{b}\frac{y_{.j}^2}{a}$$

它有 b 个自由度. 因此, 在拟合过 μ 与 $\{\beta_j\}$ 之后, $\{\tau_i\}$ 对平方和的贡献是

$$R(\tau|\mu,\beta) = R(\mu,\tau,\beta) - R(\mu,\beta) = R(\text{全模型}) - R(\text{简化模型})$$
$$= \sum_{i=1}^{a}\frac{y_{i.}^2}{b} + \sum_{j=1}^{b}\frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab} - \sum_{j=1}^{b}\frac{y_{.j}^2}{a} = \sum_{i=1}^{a}\frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab}$$

我们可以把它看作是自由度为 $a-1$ 的处理平方和 [(4.10) 式].

区组平方和可以由拟合**简化模型**

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

得出, 它也是单因子分析. 还有, 类似于 (3.5) 式, 拟合这个模型的简化平方和是

$$R(\mu,\tau) = \sum_{i=1}^{a}\frac{y_{i.}^2}{b}$$

它有 a 个自由度. 在拟合过 μ 和 $\{\tau_i\}$ 之后, 区组 $\{\beta_j\}$ 对平方和的贡献是

$$R(\beta|\mu,\tau) = R(\mu,\tau,\beta) - R(\mu,\tau)$$
$$= \sum_{i=1}^{a}\frac{y_{i.}^2}{b} + \sum_{j=1}^{b}\frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab} - \sum_{i=1}^{a}\frac{y_{i.}^2}{b} = \sum_{j=1}^{b}\frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}$$

它有 $b-1$ 个自由度, 并且它就是我们前面见到过的 (4.11) 式.

我们用一般的回归显著性检验法, 研究了随机化完全区组设计的处理、区组和误差的平方和. 虽然在随机化完全区组设计中实际分析数据时, 我们通常不用一般回归显著性检验法, 但这一方法有时在更一般的随机化区组设计中却是有用的, 正如在 4.4 节的讨论中所看到的那样.

缺失值问题的精确分析

在 4.1.3 节中, 我们给出了 RCBD 中处理缺失值的一种近似方法. 这一近似分析法由估计缺失值使误差均方值最小化所构成. 可以证明, 这一近似分析法产生一个依下述意义下处理的有偏均方值, 即, 当零假设为真时, $\mathrm{E}(MS_{\text{处理}})$ 大于 $\mathrm{E}(MS_E)$. 因此, 将得出过多的显著性结论.

缺失值问题可以用一般回归的显著性检验法来精确地分析. 缺失值会造成设计的**不平衡**, 因为并不是所有的处理都出现在所有的区组中, 我们称之为处理与区组不是*正交*的. 这一分析法还可用到类型更一般的随机化区组设计中去, 这将在 4.4 节中进一步讨论. 思考题 4.30 请读者对带有一个缺失值的随机化完全区组设计运用这一精确分析法.

4.2 拉丁方设计

4.1 节介绍了随机化完全区组设计, 这一设计可以减少由一个已知的可控制的讨厌变量的变异性所引起的实验残差. 还有很多其他类型的利用区组化原理的设计. 例如, 实验者研究用在机组人员救生系统中的火箭推进剂的 5 种不同配方关于燃烧率的效应. 每种配方都取自一批仅仅够配成 5 份供试验用的配方的原料. 而且, 这些配方是由几个操作人员准备的, 这些操作人员在实验技能和经验上可能有实质性的差别. 这样一来, 在设计中看来会有两个讨厌因子需要被 "平均出来": 原料的批次和操作人员. 这一问题的合适设计应是对每批原料的每种配方恰好试验一次, 而且对于每种配方, 5 名操作人员每人恰好试验一次. 设计的结果如表 4.8 所示, 称为**拉丁方设计** (Latin square design). 这一设计是按正方形排列的, 5 种配方 (或处理) 用拉丁字母 A, B, C, D, E 表示, 因而叫做拉丁方. 原料的批次 (行) 与操作人员 (列) 对于处理是正交的.

表 4.8 火箭推进剂问题的拉丁方设计

原料的批次	操作人员				
	1	2	3	4	5
1	$A=24$	$B=20$	$C=19$	$D=24$	$E=24$
2	$B=17$	$C=24$	$D=30$	$E=27$	$A=36$
3	$C=18$	$D=38$	$E=26$	$A=27$	$B=21$
4	$D=26$	$E=31$	$A=26$	$B=23$	$C=22$
5	$E=22$	$A=30$	$B=20$	$C=29$	$D=31$

拉丁方设计用来消除两个讨厌的变异性来源. 也就是说, 它允许从两个方向来系统地区组化. 这样一来, 行和列实际上就表示了两种随机化约束. 一般说来, 一个 p 因子的拉丁方或 $p\times p$ 拉丁方, 是一个含有 p 行和 p 列的方形. p^2 个单元中的每一个单元含有与处理对应的 p 个字母之一, 每一字母在每行每列中恰好出现一次. 一些拉丁方的例子如

4 × 4	5 × 5	6 × 6
$A\ B\ D\ C$	$A\ D\ B\ E\ C$	$A\ D\ C\ E\ B\ F$
$B\ C\ A\ D$	$D\ A\ C\ B\ E$	$B\ A\ E\ C\ F\ D$
$C\ D\ B\ A$	$C\ B\ E\ D\ A$	$C\ E\ D\ F\ A\ B$
$D\ A\ C\ B$	$B\ E\ A\ C\ D$	$D\ C\ F\ B\ E\ A$
	$E\ C\ D\ A\ B$	$F\ B\ A\ D\ C\ E$
		$E\ F\ B\ A\ D\ C$

拉丁方的**统计模型**是

$$y_{ijk}=\mu+\alpha_i+\tau_j+\beta_k+\epsilon_{ijk}\begin{cases} i=1,2,\cdots,p \\ j=1,2,\cdots,p \\ k=1,2,\cdots,p \end{cases} \tag{4.22}$$

其中 y_{ijk} 是第 j 种处理的第 i 行第 k 列的观测值, μ 是总均值, α_i 是第 i 个行效应, τ_j 是第 j 种处理效应, β_k 是第 k 个列效应, ε_{ijk} 是随机误差. 注意这是一个**效应模型**. 这一模型是完全**可加的**, 也就是说, 行、列与处理之间都没有交互作用. 因为每一单元只有一个观测值, 所以对每个

指定的观测值, 只需表示出 i, j, k 这 3 个右下标中的两个就可以了. 例如, 表 4.8 中的火箭推进剂问题, 当 i=2 且 k=3 时, 自动得 j=4(配方 D), 当 i=1 且 j=3(配方 C) 时得 k=3. 这是因为每种处理在每行每列中恰好出现一次.

方差分析是把 $N=p^2$ 个观测值的总平方和分解为行、列、处理与误差的分量, 例如

$$SS_T = SS_{\text{行}} + SS_{\text{列}} + SS_{\text{处理}} + SS_E \tag{4.23}$$

相应的自由度是

$$p^2 - 1 = p - 1 + p - 1 + p - 1 + (p-2)(p-1)$$

在通常的 ε_{ijk} 是 NID(0, σ^2) 的假设下, (4.23) 式右端项的每个平方和除以 σ^2, 就是独立分布的卡方随机变量. 用来检验处理均值没有差异的合适统计量是

$$F_0 = \frac{MS_{\text{处理}}}{MS_E}$$

在零假设下其分布为 $F_{p-1,(p-2)(p-1)}$. 我们也可以检验没有行效应与没有列效应, 只要用 $MS_{\text{行}}$(或 $MS_{\text{列}}$) 与 MS_E 的比值即可. 不过, 由于行和列表示随机化约束, 这些检验可能未必合适.

方差分析计算方法中的处理、行以及列之和如表 4.9 所示. 从平方和的计算公式可以看出, 这一分析只是随机化完全区组设计的简单延伸, 其中源于行因子的平方和是用那些行和得到的.

表 4.9　拉丁方设计的方差分析

方差来源	平方和	自由度	均方	F_0
处理	$SS_{\text{处理}} = \frac{1}{p}\sum_{j=1}^{p} y_{.j.}^2 - \frac{y_{...}^2}{N}$	$p-1$	$\frac{SS_{\text{处理}}}{p-1}$	$F_0 = \frac{MS_{\text{处理}}}{MS_E}$
行	$SS_{\text{行}} = \frac{1}{p}\sum_{i=1}^{p} y_{i..}^2 - \frac{y_{...}^2}{N}$	$p-1$	$\frac{SS_{\text{行}}}{p-1}$	
列	$SS_{\text{列}} = \frac{1}{p}\sum_{k=1}^{p} y_{..k}^2 - \frac{y_{...}^2}{N}$	$p-1$	$\frac{SS_{\text{列}}}{p-1}$	
误差	SS_E(用减法)	$(p-2)(p-1)$	$\frac{SS_E}{(p-2)(p-1)}$	
总和	$SS_T = \sum_i\sum_j\sum_k y_{ijk}^2 - \frac{y_{...}^2}{N}$	P^2-1		

例 4.3 考虑前面的火箭推进剂问题, 其中原料的批次与操作人员都代表随机化约束. 这一实验设计如表 4.8 所示, 是一个 5×5 拉丁方.

表 4.10　火箭推进剂问题的规范数据

原料的批次	操作人员					$y_{i..}$
	1	2	3	4	5	
1	$A=-1$	$B=-5$	$C=-6$	$D=-1$	$E=-1$	-14
2	$B=-8$	$C=-1$	$D=5$	$E=2$	$A=11$	9
3	$C=-7$	$D=13$	$E=1$	$A=2$	$B=-4$	5
4	$D=1$	$E=6$	$A=1$	$B=-2$	$C=-3$	3
5	$E=-3$	$A=5$	$B=-5$	$C=4$	$D=6$	7
$y_{..k}$	-18	18	-4	5	9	$10=y_{...}$

现将每个观测值减 25 得如表 4.10 所示的规范值. 总平方和、批次 (行) 平方和、操作人员 (列) 平方和算得如下:

$$SS_T = \sum_i\sum_j\sum_k y_{ijk}^2 - \frac{y_{...}^2}{N} = 680 - \frac{(10)^2}{25} = 676.00$$

$$SS_{\text{批次}} = \frac{1}{p}\sum_{i=1}^{p} y_{i..}^2 - \frac{y_{...}^2}{N} = \frac{1}{5}[(-14)^2 + 9^2 + 5^2 + 3^2 + 7^2] - \frac{(10)^2}{25} = 68.00$$

$$SS_{\text{操作人员}} = \frac{1}{p}\sum_{k=1}^{p} y_{..k}^2 - \frac{y_{...}^2}{N} = \frac{1}{5}[(-18)^2 + 18^2 + (-4)^2 + 5^2 + 9^2] - \frac{(10)^2}{25} = 150.00$$

处理 (拉丁字母) 的总和是

拉丁字母	A	B	C	D	E
处理总和	$y_{.1.} = 18$	$y_{.2.} = -24$	$y_{.3.} = -13$	$y_{.4.} = 24$	$y_{.5.} = 5$

由这些总和算得配方的平方和为

$$SS_{\text{配方}} = \frac{1}{p}\sum_{j=1}^{p} y_{.j.}^2 - \frac{y_{...}^2}{N} = \frac{1}{5}[18^2 + (-24)^2 + (-13)^2 + 24^2 + 5^2] - \frac{(10)^2}{25} = 330.00$$

相减得误差的平方和：

$$SS_E = SS_T - SS_{\text{批次}} - SS_{\text{操作人员}} - SS_{\text{配方}} = 676.00 - 68.00 - 150.00 - 330.00 = 128.00$$

方差分析见表 4.11. 结论是，不同的火箭推进剂配方所产生的燃烧率有显著性差异. 方差分析还表明，操作人员之间亦有差异，因此，关于该因子的区组化的确是一个好的预防措施. 没有强有力的证据证明原料的批次间存在差异，所以，在这一实验中，看来不必关心这一变异性来源. 不过，对原料的批次区组化通常是一个好的主意.

表 4.11 火箭推进剂配方问题的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
配方	330.00	4	82.50	7.73	0.002 5
原料的批次	68.00	4	17.00		
操作人员	150.00	4	37.50		
误差	128.00	12	10.67		
总和	676.00	24			

如同在任一设计问题中一样，实验者应该通过检查并绘制残差图来考察模型的适合性. 对于某一拉丁方，其残差为

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk} = y_{ijk} - \overline{y}_{i..} - \overline{y}_{.j.} - \overline{y}_{..k} + 2\overline{y}_{...}$$

请读者计算例 4.3 的残差并绘制适当的图.

第一行和第一列按字母顺序构成的拉丁方叫做**标准拉丁方** (standard Latin square). 例 4.4 所用的设计就是一种标准拉丁方. 标准拉丁方通常可将第一行按字母顺序写，然后后续的每一行都是将上一行的字母向左移一个位置. 表 4.12 概括了关于拉丁方及标准拉丁方的一些重要事实.

与任何实验设计一样，拉丁方中的观测值也应以随机顺序取得. 常用的随机化方法是随机地选取一特定的拉丁方来使用. 正如在表 4.12 中所见到的，一个指定大小的拉丁方有很多个，因此，不可能枚举出所有的拉丁方并随机地选择一个. 普遍的方法是在如 Fisher and Yates(1953) 所设计的那样的表中选取一个拉丁方. 然后按行、列以及字母随机地排列. 这个问题在 Fisher and Yates(1953) 的书中有更全面的讨论.

有时，拉丁方中的一个观测值会缺失，对于一个 $p \times p$ 拉丁方，这一缺失值可以估计为

$$y_{ijk} = \frac{p(y'_{i..} + y'_{.j.} + y'_{..k}) - 2y'_{...}}{(p-2)(p-1)} \tag{4.24}$$

表 4.12 标准拉丁方和不同大小拉丁方的数量 [a]

大 小	3×3	4×4	5×5	6×6	7×7	$p \times p$
标准方的数量	ABC	$ABCD$	$ABCDE$	$ABCDEF$	$ABCDEFG$	$ABC...P$
	BCA	$BCDA$	$BAECD$	$BCFADE$	$BCDEFGA$	$BCD...A$
	CAB	$CDAB$	$CDAEB$	$CFBEAD$	$CDEFGAB$	$CDE...B$
		$DABC$	$DEBAC$	$DEABFC$	$DEFGABC$	⋮
			$ECDBA$	$EADFCB$	$EFGABCD$	
				$FDECBA$	$FGABCDE$	$PAB...(P-1)$
					$GABCDEF$	
标准方的例子	1	4	56	9 408	16 942 080	—
标准方的数目	12	576	161 280	818 851 200	61 479 419 904 000	$p!(p-1)!\times$ 拉丁方的总数

a 该表的一些信息见*Statistical Tables for Biological, Agricultural and Medical Research*, 4th edition, by R. A. Fisher and F. Yates, Oliver and Boyd, Edinburgh, 1953. 人们对于超过 7×7 的拉丁方的特性了解较少.

其中, 撇号表示有缺失值的行、列和处理的总和, $y'_{..}$ 是有缺失值时的全体观测值的总和.

拉丁方可用于行和列代表实验者实际想要研究的因子的情况, 其中没有随机化约束. 这样一来, 3 个因子 (行、列、字母), 每一个有 p 个水平, 能够仅用 p^2 次试验来研究. 这种设计假定因子之间没有交互作用. 更详尽的内容随后将在交互作用的论题中讨论.

1. 拉丁方的重复

小的拉丁方的一个缺点是它给出相对小的误差自由度. 例如, 一个 3×3 拉丁方只有 2 个误差自由度, 一个 4×4 拉丁方只有 6 个误差自由度, 等等. 当使用小的拉丁方时, 常常希望重复使用它们以便增加误差自由度.

重复拉丁方有几种方法. 为说明起见, 考虑例 4.4 所用的 5×5 拉丁方, 假定它重复 n 次. 做法有如下几种.

(1) 每次重复中用相同的批次和相同的操作人员.

(2) 每次重复中用相同的批次但用不同的操作人员 (或者, 等价地, 用相同的操作人员但用不同的批次).

(3) 用不同的批次和不同的操作人员.

方差分析将依赖于重复的方法.

考虑情形 1, 此时用于每次重复的行和列区组化因子的水平相同. 令 y_{ijkl} 表示行 i、处理 j、列 k、第 l 次重复的观测值, 有 $N=np^2$ 个总的观测值. 方差分析法概括在表 4.13 中.

现考虑情形 2, 设在每次重复中用新的原料批次但操作人员不变. 这样一来, 在每次重复中有 5 个新的行 (一般地, 有 p 个新行). 方差分析概括在表 4.14 中. 注意, 行的方差来源实际上度量了在 n 次重复中行间的变异.

最后, 考虑情形 3, 此时在每次重复中用新的原料批次和新的操作人员. 现在, 行和列所得的方差度量了在重复中相应因子的变异. 方差分析法概括在表 4.15 中.

分析重复的拉丁方还有其他方法, 它容许处理与拉丁方之间有某种交互作用 (参见思考题

4.23).

表 4.13　重复拉丁方的方差分析, 情形 1

方差来源	平方和	自由度	均方	F_0
处理	$\frac{1}{np}\sum_{j=1}^{p} y_{.i..}^2 - \frac{y_{....}^2}{N}$	$p-1$	$\frac{SS_{处理}}{p-1}$	$\frac{MS_{处理}}{MS_E}$
行	$\frac{1}{np}\sum_{i=1}^{p} y_{i...}^2 - \frac{y_{....}^2}{N}$	$p-1$	$\frac{SS_{行}}{p-1}$	
列	$\frac{1}{np}\sum_{k=1}^{p} y_{..k.}^2 - \frac{y_{....}^2}{N}$	$p-1$	$\frac{SS_{列}}{p-1}$	
重复	$\frac{1}{p^2}\sum_{l=1}^{n} y_{...l}^2 - \frac{y_{....}^2}{N}$	$n-1$	$\frac{SS_{重复}}{n-1}$	
误差	用减法	$(p-1)[n(p+1)-3]$	$\frac{SS_E}{(p-1)[n(p+1)-3]}$	
总和	$\sum\sum\sum\sum y_{ijkl}^2 - \frac{y_{....}^2}{N}$	np^2-1		

表 4.14　重复拉丁方的方差分析, 情形 2

方差来源	平方和	自由度	均方	F_0
处理	$\frac{1}{np}\sum_{j=1}^{p} y_{.j..}^2 - \frac{y_{....}^2}{N}$	$p-1$	$\frac{SS_{处理}}{p-1}$	$\frac{MS_{处理}}{MS_E}$
行	$\frac{1}{p}\sum_{l=1}^{n}\sum_{j=1}^{p} y_{i..l}^2 - \sum_{l=1}^{n}\frac{y_{...l}^2}{N}$	$n(p-1)$	$\frac{SS_{行}}{n(p-1)}$	
列	$\frac{1}{np}\sum_{k=1}^{p} y_{..k.}^2 - \frac{y_{....}^2}{N}$	$p-1$	$\frac{SS_{列}}{p-1}$	
重复	$\frac{1}{p^2}\sum_{l=1}^{n} y_{...l}^2 - \frac{y_{....}^2}{N}$	$n-1$	$\frac{SS_{重复}}{n-1}$	
误差	用减法	$(p-1)(np-1)$	$\frac{SS_E}{(p-1)(np-1)}$	
总和	$\sum_i\sum_j\sum_k\sum_l y_{ijkl}^2 - \frac{y_{....}^2}{N}$	np^2-1		

表 4.15　重复拉丁方的方差分析, 情形 3

方差来源	平方和	自由度	均方	F_0
处理	$\frac{1}{np}\sum_{j=1}^{p} y_{.j..}^2 - \frac{y_{....}^2}{N}$	$p-1$	$\frac{SS_{处理}}{p-1}$	$\frac{MS_{处理}}{MS_E}$
行	$\frac{1}{p}\sum_{l=1}^{n}\sum_{i=1}^{p} y_{i..l}^2 - \sum_{l=1}^{n}\frac{y_{...l}^2}{p^2}$	$n(p-1)$	$\frac{SS_{行}}{n(p-1)}$	
列	$\frac{1}{p}\sum_{l=1}^{n}\sum_{k=1}^{p} y_{..kl}^2 - \sum_{l=1}^{n}\frac{y_{...l}^2}{p^2}$	$n(p-1)$	$\frac{SS_{列}}{n(p-1)}$	
重复	$\frac{1}{p^2}\sum_{l=1}^{n} y_{...l}^2 - \frac{y_{....}^2}{N}$	$n-1$	$\frac{SS_{重复}}{n-1}$	
误差	用减法	$(p-1)[n(p-1)-1]$	$\frac{SS_E}{(p-1)[n(p-1)-1]}$	
总和	$\sum_i\sum_j\sum_k\sum_l y_{ijkl}^2 - \frac{y_{....}^2}{N}$	np^2-1		

2. 交叉设计和平衡剩余效应设计

有时, 人们会遇到时间周期是实验的一个因子这样的问题. 一般说来, 在 p 个时间周期内要检验 p 种处理时用 np 个实验单元. 例如, 一位人类行为分析者研究 20 个对象中两种流质的脱水效应. 在第一个周期内, 有一半对象 (随机选取) 给流质 A, 另一半对象给流质 B. 在周期之末, 测出响应值, 并且可以忽略某个周期以消除流质的任何生理效应. 实验者再使原来取流质 A 的对象取流质 B, 原来取流质 B 的对象取流质 A. 这样的设计叫做*交叉设计*. 它可以作为有两行 (时间周期) 和两种处理 (流质类型) 的 10 个拉丁方组来进行分析. 10 个拉丁方的每一个中的两列都与对象相对应.

这一设计的布局如图 4.7 所示. 拉丁方中的行代表时间周期, 列代表对象. 首先接受流质 A 的 10 个对象 (1, 4, 6, 7, 9, 12, 13, 15, 17, 19) 是随机确定的.

拉丁方

	I		II		III		IV		V		VI		VII		VIII		IX		X	
对象	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
周期 1	A	B	B	A	B	A	A	B	A	B	B	A	A	B	A	B	A	B	A	B
周期 2	B	A	A	B	A	B	B	A	B	A	A	B	B	A	B	A	B	A	B	A

图 4.7 交叉设计

简略的方差分析概括在表 4.16 中. 对象平方和作为 20 个对象总数之间的校正平方和来计算, 周期平方和是行间的校正平方和, 流质平方和作为字母总数间的校正平方和来计算. 这些设计的更详细的统计分析可参阅 Cochran and Cox(1957), John(1971), 以及 Anderson and McLean(1974).

我们也可以将拉丁方型设计用于处理有剩余效应的实验中去, 也就是说, 例如, 在周期 2 中流质 B 的数据仍然反映出周期 1 中流质 A 的一些效应. Cochran and Cox(1957) 以及 John(1971) 对有剩余效应的平衡设计进行了详细的讨论.

表 4.16 图 4.7 中的交叉设计的方差分析

方差来源	自由度
对象 (列)	19
周期 (行)	1
流质 (字母)	1
误差	18
总和	39

4.3 正交拉丁方设计

考虑一个 $p \times p$ 拉丁方, 在其上叠加第 2 个 $p \times p$ 拉丁方, 其处理用希腊字母表示. 如果在叠加时使每个希腊字母与拉丁字母一起出现一次且仅有一次, 则称两个拉丁方是*正交的*, 所得的

设计叫做**正交拉丁方**[①]. 一个 4×4 正交拉丁方的例子如表 4.17 所示.

表 4.17 4×4 正交拉丁方设计

行	列			
	1	2	3	4
1	$A\alpha$	$B\beta$	$C\gamma$	$D\delta$
2	$B\delta$	$A\gamma$	$D\beta$	$C\alpha$
3	$C\beta$	$D\alpha$	$A\delta$	$B\gamma$
4	$D\gamma$	$C\delta$	$B\alpha$	$A\beta$

正交拉丁方设计能够用来系统地控制 3 个外部变异性的来源, 也就是说, 可以沿 3 个方向划分区组. 这一设计可以考察 4 个因子 (行、列、拉丁字母、希腊字母), 每个因子有 p 个水平, 仅需作 p^2 次试验. 正交拉丁方除了 $p=6$ 之外对所有的 $p\geqslant 3$ 都存在.

正交拉丁方设计的统计模型是

$$y_{ijkl}=\mu+\theta_i+\tau_j+\omega_k+\Psi_l+\epsilon_{ijkl}\begin{cases} i=1,2,\cdots,p\\ j=1,2,\cdots,p\\ k=1,2,\cdots,p\\ l=1,2,\cdots,p\end{cases} \tag{4.25}$$

其中 y_{ijkl} 是对拉丁字母 j 和希腊字母 k 在行 i 和列 l 上的观测值, θ_i 是第 i 行的效应, τ_j 是拉丁字母处理 j 的效应, ω_k 是希腊字母处理 k 的效应, ψ_l 是列 l 的效应, ϵ_{ijkl} 是一个服从 NID(0, σ^2) 的随机误差分量. 完全识别一个观测值只需要 4 个下标中的 2 个.

方差分析与拉丁方的分析非常相像. 因为希腊字母在每一行和每一列中恰出现一次, 与每一拉丁字母一起也恰出现一次, 由希腊字母代表的因子正交于行、列以及拉丁字母表示的处理. 因此, 希腊字母因子的平方和可以由希腊字母的总和来计算, 而且实验误差可以进一步用这一问题来减小. 计算上的细节由表 4.18 来说明. 行、列、拉丁字母和希腊字母处理相等的零假设可以用对应的均方值除以均方误差来检验. 拒绝域是 $F_{p-1,(p-3)(p-1)}$ 分布的上尾部点.

表 4.18 正交拉丁方设计的方差分析

方差来源	平方和	自由度
拉丁字母处理	$SS_L=\frac{1}{p}\sum\limits_{j=1}^{p}y_{.j..}^2-\frac{y_{....}^2}{N}$	$p-1$
正交字母处理	$SS_G=\frac{1}{p}\sum\limits_{k=1}^{p}y_{..k.}^2-\frac{y_{....}^2}{N}$	$p-1$
行	$SS_{行}=\frac{1}{p}\sum\limits_{i=1}^{p}y_{i...}^2-\frac{y_{....}^2}{N}$	$p-1$
列	$SS_{列}=\frac{1}{p}\sum\limits_{l=1}^{p}y_{...l}^2-\frac{y_{....}^2}{N}$	$p-1$
误差	SS_E(用减法)	$(p-3)\ (p-1)$
总和	$SS_T=\sum\limits_i\sum\limits_j\sum\limits_k\sum\limits_l y_{ijkl}^2-\frac{y_{....}^2}{N}$	p^2-1

① 也有的书称之为希腊拉丁方. —— 译者注

例 4.4 设在例 4.3 的火箭推进剂实验中, 还需考虑一个可能重要的因子: 试验装配. 设有 5 种试验装配, 用希腊字母 α, β, γ, δ, ε 表示. 5×5 的正交拉丁方设计的结果如表 4.19 所示.

表 4.19 火箭推进剂实验的 5×5 的正交拉丁方设计

原料的批次	操作人员					$y_{i..}$
	1	2	3	4	5	
1	$A\alpha=-1$	$B\gamma=-5$	$C\epsilon=-6$	$D\beta=-1$	$E\delta=-1$	−14
2	$B\beta=-8$	$C\delta=-1$	$D\alpha=5$	$E\gamma=2$	$A\epsilon=11$	9
3	$C\gamma=-7$	$D\epsilon=13$	$E\beta=1$	$A\delta=2$	$B\alpha=-4$	5
4	$D\delta=1$	$E\alpha=6$	$A\gamma=1$	$B\epsilon=-2$	$C\beta=-3$	3
5	$E\epsilon=-3$	$A\beta=5$	$B\delta=-5$	$C\alpha=4$	$D\gamma=6$	7
$y_{..l}$	−18	18	−4	5	9	$10=y_{....}$

因为原料的批次 (行)、操作人员 (列)、配方 (拉丁字母) 的总和与例 4.3 的相等, 所以有

$$SS_{\text{批次}}=68.00\quad SS_{\text{操作人员}}=150.00\quad SS_{\text{配方}}=330.00$$

试验装配 (希腊字母) 的总和是

希腊字母	试验装配总和
α	$y_{..1.}=10$
β	$y_{..2.}=-6$
γ	$y_{..3.}=-3$
δ	$y_{..4.}=-4$
ϵ	$y_{..5.}=13$

于是, 试验装配的平方和是

$$SS_{\text{装配}}=\frac{1}{p}\sum_{k=1}^{p}y_{..k.}^2-\frac{y_{....}^2}{N}=\frac{1}{5}[10^2+(-6)^2+(-3)^2+(-4)^2+13^2]-\frac{(10)^2}{25}=62.00$$

完全的方差分析概括在表 4.20 中. 配方的显著性差异为 1%. 将表 4.20 与表 4.11 进行比较可以看出, 移走试验装配的变异性会减少实验误差. 然而, 在减少实验误差时, 也将误差自由度从 12(例 4.3 的拉丁方设计) 减少至 8. 这样一来, 误差的估计量少了几个自由度, 从而可能使检验降低灵敏度.

表 4.20 火箭推进剂配方问题的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
配方	330.00	4	82.50	10.00	0.003 3
原料的批次	68.00	4	17.00		
操作人员	150.00	4	37.50		
试验装配	62.00	4	15.50		
误差	66.00	8	8.25		
总和	676.00	24			

构成正交拉丁方的拉丁方正交对的概念可以作些推广. 一个 $p\times p$ **超方** (hypersquare) 是一种设计, 其中叠加有 3 个或更多的互相正交的 $p\times p$ 拉丁方. 一般说来, 如果能得到一个有 $p-1$ 个互相正交的拉丁方的完全组, 就可以研究多至 $p+1$ 个因子. 这种设计将利用所有 $(p+1)(p-1)=p^2-1$ 个自由度, 所以, 一个独立的误差方差的估计是必需的. 当然, 在用超方时, 因子之间必须没有交互作用.

4.4 平衡不完全区组设计

在使用随机化区组设计的某些实验中, 对于每一区组, 我们未必都能够对所有的处理组合各

做一次试验. 这种情况的发生通常是因为缺乏实验仪器或器材, 或者是由于区组的实际大小等原因. 例如, 在血管移植实验中 (例 4.1), 假定每一批原材料的大小仅可供检验 3 个挤压压强. 因此, 每个挤压压强就不能在每批原材料都得到检验. 对于这类问题还是有可能来使用随机化区组设计的, 不过, 不是每个处理都在每个区组中出现. 这种设计叫做**随机化不完全区组设计** (randomized incomplete block design).

当所有的处理比较具有同等的重要性时, 用于每一区组的处理组合应该按平衡方式来选取, 也就是说, 使任意一对处理在一起出现的次数相同. **平衡不完全区组设计**(BIBD) 就是这样一种不完全区组设计, 其中, 任意两个处理在一起出现的次数相等. 假定有 a 个处理, 在每一区组中刚好容纳 $k(k < a)$ 个处理. 平衡不完全区组设计可以取 $\binom{a}{k}$ 个区组并对每一区组安排一个不同的处理组合. 不过, 常常可以取少于 $\binom{a}{k}$ 个区组来得到平衡的设计. Fisher and Yates(1953), Davies(1956) 与 Cochran and Cox(1957) 都给出了 BIBD 表.

例如, 设一位化学工程师将某一化学过程的反应时间看作为所用的催化剂类型的函数. 现在研究 4 种催化剂. 实验方法是, 选取一批原材料装入实验设备中, 将每一种催化剂分别在该实验设备中进行试验, 观测反应时间. 因为每批原材料的变化可能会影响到催化剂的作用, 工程师决定将原材料的批次看作为区组. 但是, 每批原材料只够供试验 3 种催化剂. 因此, 必须用随机化不完全区组设计. 这一实验的平衡不完全区组设计, 以及观测记录如表 4.21 所示. 催化剂在每一区组中实验的顺序是随机的.

表 4.21 催化剂试验的平衡不完全区组设计

处理 (催化剂)	区组 (原料的批次)				
	1	2	3	4	$y_{i.}$
1	73	74	—	71	218
2	—	75	67	72	214
3	73	75	68	—	216
4	75	—	72	75	222
$y_{.j}$	221	224	207	218	$870=y_{..}$

4.4.1 BIBD 的统计分析

与通常一样, 我们假定有 a 种处理和 b 个区组. 此外, 还假定每一区组含有 k 种处理, 每一处理在设计中出现 r 次 (或者说, 重复 r 次), 从而有 $N= ar = bk$ 个总的观测值. 还有, 每对处理在同一区组中出现的次数是

$$\lambda = \frac{r(k-1)}{a-1}$$

当 $a = b$ 时, 称这一设计是**对称的**.

参数 λ 必须是一个整数. 为导出关于 λ 的关系式, 我们考虑任一处理, 比方说处理 1. 因为处理 1 出现在 r 个区组中, 而在这类区组的每一个区组中都有 $k-1$ 个其余的处理, 因此有 $r(k-1)$ 个观测值包含在处理 1 的区组中. 这 $r(k-1)$ 个观测值也必须表示其余 $a-1$ 个处理 λ 次. 因此 $\lambda(a-1) = r(k-1)$.

BIBD 的统计模型是

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \tag{4.26}$$

其中 y_{ij} 是第 j 个区组的第 i 个观测值, μ 是总均值, τ_i 是第 i 种处理的效应, β_j 是第 j 个区组的效应, ϵ_{ij} 是 NID(0, σ^2) 随机误差分量. 数据的总方差是总校正平方和:

$$SS_T = \sum_i \sum_j y_{ij}^2 - \frac{y_{..}^2}{N} \tag{4.27}$$

总变异性可分解为

$$SS_T = SS_{\text{处理 (已调整的)}} + SS_{\text{区组}} + SS_E$$

其中处理平方和已被调整, 使处理与区组效应得到分离. 这一调整是必需的, 因为每一处理所在的 r 个区组集是相异的. 这样一来, 未曾调整的处理总和 $y_{1.}, y_{2.}, \cdots, y_{a.}$ 之间的差亦受区组之间差的影响.

区组平方和是

$$SS_{\text{区组}} = \frac{1}{k}\sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{N} \tag{4.28}$$

其中 $y_{.j}$ 是第 j 个区组的总和, $SS_{\text{区组}}$ 有 $b-1$ 个自由度. 已调整的处理平方和是

$$SS_{\text{处理 (已调整的)}} = \frac{k\sum_{i=1}^{a} Q_i^2}{\lambda a} \tag{4.29}$$

其中 Q_i 是第 i 种处理的已调整的总和, 其计算式为

$$Q_i = y_{i.} - \frac{1}{k}\sum_{j=1}^{b} n_{ij} y_{.j} \quad i = 1, 2, \cdots, a \tag{4.30}$$

当区组 j 中出现处理 i 时 $n_{ij} = 1$, 否则 $n_{ij} = 0$. 已调整的处理总和加起来通常为零. $SS_{\text{处理 (已调整的)}}$ 有 $a-1$ 个自由度. 由减法算得误差平方和为

$$SS_E = SS_T - SS_{\text{处理 (已调整的)}} - SS_{\text{区组}} \tag{4.31}$$

它有 $N-a-b+1$ 个自由度.

检验处理效应相等的恰当的统计量是

$$F_0 = \frac{MS_{\text{处理 (已调整的)}}}{MS_E}$$

方差分析概括在表 4.22 中.

表 4.22 平衡不完全区组设计的方差分析

方差来源	平方和	自由度	均　方	F_0
处理 (已调整的)	$\frac{k\sum Q_i^2}{\lambda a}$	$a-1$	$\frac{SS_{\text{处理 (已调整的)}}}{a-1}$	$F_0 = \frac{MS_{\text{处理 (已调整)}}}{MS_E}$
区组	$\frac{1}{k}\sum y_{.j}^2 - \frac{y_{..}^2}{N}$	$b-1$	$\frac{SS_{\text{区组}}}{b-1}$	
误差	SS_E(用减法)	$N-a-b+1$	$\frac{SS_E}{N-a-b+1}$	
总和	$\sum_i \sum_j y_{ij}^2 - \frac{y_{..}^2}{N}$	$N-1$		

例 4.5 考虑表 4.21 中关于催化剂实验的数据. 这是 $a=4, b=4, k=3, r=3, \lambda=2$ 以及 $N=12$ 的 BIBD. 数据的分析如下. 总平方和是

$$SS_T = \sum_i \sum_j y_{ij}^2 - \frac{y_{..}^2}{12} = 63\ 156 - \frac{(870)^2}{12} = 81.00$$

由 (4.28) 式得区组平方和为

$$SS_{区组} = \frac{1}{3}\sum_{j=1}^{4} y_{.j}^2 - \frac{y_{..}^2}{12} = \frac{1}{3}[(221)^2 + (207)^2 + (224)^2 + (218)^2] - \frac{(870)^2}{12} = 55.00$$

要计算对区组已调整的处理平方和, 先用 (4.30) 式确定出已调整的处理总和为

$$Q_1 = (218) - \tfrac{1}{3}(221 + 224 + 218) = -9/3$$
$$Q_2 = (214) - \tfrac{1}{3}(207 + 224 + 218) = -7/3$$
$$Q_3 = (216) - \tfrac{1}{3}(221 + 207 + 224) = -4/3$$
$$Q_4 = (222) - \tfrac{1}{3}(221 + 207 + 218) = 20/3$$

再由 (4.29) 式算得已调整的处理平方和为

$$SS_{处理\ (已调整的)} = \frac{k\sum_{i=1}^{4} Q_i^2}{\lambda a} = \frac{3[(-9/3)^2 + (-7/3)^2 + (-4/3)^2 + (20/3)^2]}{(2)(4)} = 22.75$$

相减得误差平方和为

$$SS_E = SS_T - SS_{处理\ (已调整的)} - SS_{区组} = 81.00 - 22.75 - 55.00 = 3.25$$

方差分析如表 4.23 所示. 因为 p 值较小, 所以结论是, 所用的催化剂对反应时间有显著效应.

表 4.23 例 4.5 的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
处理 (对区组调整的)	22.75	3	7.58	11.66	0.010 7
区组	55.00	3	—		
误差	3.25	5	0.65		
总和	81.00	11			

如果所研究的因子是固定的, 则对各个处理均值进行检验是有意思的. 如果使用正交对照的话, 则这些对照一定是由**已调整的处理总和**即 $\{Q_i\}$ 而不是由 $\{y_{i.}\}$ 所得出. 对照平方和是

$$SS_C = \frac{k\left(\sum_{i=1}^{a} c_i Q_i\right)^2}{\lambda a \sum_{i=1}^{a} c_i^2}$$

其中 $\{c_i\}$ 是对照系数. 其他的多重比较法亦可以用来比较所有已调整过的处理均值对, 在 4.4.2 节中我们会见到它可以用 $\hat{\tau}_i = kQ_i/(\lambda a)$ 来估计. 已调整的处理效应的标准误是

$$S = \sqrt{\frac{kMS_E}{\lambda a}} \tag{4.32}$$

上述分析中, 我们是将总平方和分解为已调整的处理平方和、未调整的区组平方和以及误差平方和. 有时, 我们想要评估区组效应, 要做到这一点, 需要 SS_T 的另一分解式, 那就是

$$SS_T = SS_{处理} + SS_{区组\ (已调整的)} + SS_E$$

此处 $SS_{处理}$ 是未调整的. 如果设计是对称的, 即 $a = b$, 则可得出 $SS_{区组（已调整的）}$ 的一个简单公式. 已调整的区组总和是

$$Q'_j = y_{\cdot j} - \frac{1}{4}\sum_{i=1}^{a} n_{ij} y_{i\cdot} \quad j = 1, 2, \cdots, b \tag{4.33}$$

从而

$$SS_{区组（已调整的）} = \frac{r\sum_{j=1}^{b}(Q'_j)^2}{\lambda b} \tag{4.34}$$

因为 $a = b = 4$, 例 4.5 的平衡不完全区组设计是对称的. 因此

$$Q'_1 = (221) - \tfrac{1}{3}(218 + 216 + 222) = 7/3$$
$$Q'_2 = (224) - \tfrac{1}{3}(218 + 214 + 216) = 24/3$$
$$Q'_3 = (207) - \tfrac{1}{3}(214 + 216 + 222) = -31/3$$
$$Q'_4 = (218) - \tfrac{1}{3}(218 + 214 + 222) = 0$$

从而

$$SS_{区组（已调整的）} = \frac{3[(7/3)^2 + (24/3)^2 + (-31/3)^2 + (0)^2]}{(2)(4)} = 66.08$$

还有

$$SS_{处理} = \frac{(218)^2 + (214)^2 + (216)^2 + (222)^2}{3} - \frac{(870)^2}{12} = 11.67$$

对称的平衡不完全区组设计的方差分析概括在表 4.24 中. 注意, 表 4.24 中与均方有关的平方和加起来不等于总平方和, 即

$$SS_T \neq SS_E + SS_{处理（已调整的）} + SS_{区组（已调整的）}$$

这是由于处理与区组的非正交性所致.

表 4.24 例 4.5 的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
处理（已调整的）	22.75	3	7.58	11.66	0.010 7
处理（未调整的）	11.67	3			
区组（未调整的）	55.00	3			
区组（调整的）	66.08	3	22.03	33.90	0.001 0
误差	3.25	5	0.65		
总和	81.00	11			

计算机输出

有许多计算机软件包可以进行平衡不完全区组设计的分析. SAS 的一般线性模型方法是其中之一, 另外一种是 Minitab, 它是 PC 广泛使用的统计软件包. 表 4.25 的上半部是例 4.5 的

Minitab 的一般线性模型. 比较表 4.25 和表 4.24, Minitab 已经计算出了已调整的处理平方和以及已调整的区组平方和. (它们在 Minitab 输出中被称为 "Adj SS").

表 4.25 例 4.5 的 Minitab(一般线性模型) 分析

```
General Linear Model
Factor       Type     Levels     Values
Catalyst     fixed         4      1 2 3 4
Block        fixed         4      1 2 3 4

Analysis of Variance for Time, using Adjusted SS for Tests
Source      DF     Seq SS    Adj SS     Adj MS        F        P
Catalyst     3     11.667    22.750      7.583    11.67    0.011
Block        3     66.083    66.083     22.028    33.89    0.001
Error        5      3.250     3.250      0.650
Total       11     81.000

Tukey 95.0% Simultaneous Confidence Intervals
Response Variable Time
All Pairwise Comparisons among Levels of Catalyst

Catalyst = 1 subtracted from:

Catalyst    Lower   Center   Upper   ---------+--------+--------+------
2          -2.327   0.2500   2.827    (--------*--------)
3          -1.952   0.6250   3.202     (--------*--------)
4           1.048   3.6250   6.202              (--------*--------)
                                     ---------+--------+--------+------
                                             0.0      2.5      5.0

Catalyst = 2 subtracted from:

Catalyst    Lower   Center   Upper   ---------+--------+--------+------
3          -2.202   0.3750   2.952    (--------*--------)
4           0.798   3.3750   5.952             (--------*--------)
                                     ---------+--------+--------+------
                                             0.0      2.5      5.0

Catalyst = 3 subtracted from:

Catalyst    Lower   Center   Upper   ---------+--------+--------+------
4          0.4228    3.000   5.577             (--------*--------)
                                     ---------+--------+--------+------
                                             0.0      2.5      5.0

Tukey Simultaneous Tests
Response Variable Time
All Pairwise Comparisons among Levels of Catalyst

Catalyst = 1 subtracted from:

Level       Difference        SE of                 Adjusted
Catalyst      of Means   Difference    T-Value       P-Value
2               0.2500       0.6982     0.3581        0.9825
3               0.6250       0.6982     0.8951        0.8085
4               3.6250       0.6982     5.1918        0.0130

Catalyst = 2 subtracted from:

Level       Difference        SE of                 Adjusted
Catalyst      of Means   Difference    T-Value       P-Value
3               0.3750       0.6982     0.5371        0.9462
4               3.3750       0.6982     4.8338        0.0175

Catalyst = 3 subtracted from:

Level       Difference        SE of                 Adjusted
Catalyst      of Means   Difference    T-Value       P-Value
4                3.000       0.6982      4.297        0.0281
```

表 4.25 的下半部是一个多重比较分析, 采用 Tukey 方法. 它列出了所有成对均值的差异的

置信区间和 Tukey 检验. 使用 Tukey 方法, 我们得出结论: 催化剂 4 与其他 3 种催化剂有差异.

4.4.2 参数的最小二乘估计

考虑估计平衡不完全区组模型的处理效应. 最小二乘正规方程组是

$$\begin{aligned}
\mu &: N\hat{\mu} + r\sum_{i=1}^{a}\hat{\tau}_i + k\sum_{j=1}^{b}\hat{\beta}_j = y_{..} \\
\tau_i &: r\hat{\mu} + r\hat{\tau}_i + \sum_{j=1}^{b} n_{ij}\hat{\beta}_j = y_{i.} \quad i = 1, 2, \cdots, a \\
\beta_j &: k\hat{\mu} + \sum_{i=1}^{a} n_{ij}\hat{\tau}_i + k\hat{\beta}_j = y_{.j} \quad j = 1, 2, \cdots, a
\end{aligned} \tag{4.35}$$

添加约束 $\sum\hat{\tau}_i = \sum\hat{\beta}_j = 0$, 可得 $\hat{\mu} = \overline{y}_{..}$. 用 $\{\beta_j\}$ 的方程组消去 $\{\tau_i\}$ 的方程组中的区组效应, 得

$$rk\hat{\tau}_i - r\hat{\tau}_i - \sum_{j=1}^{b}\sum_{\substack{p=1 \\ p\neq 1}}^{a} n_{ij}n_{pj}\hat{\tau}_p = ky_{i.} - \sum_{j=1}^{b} n_{ij}y_{.j} \tag{4.36}$$

注意 (4.36) 式的右端项就是 kQ_i, 其中 Q_i 是第 i 种已调整的处理总和 [见 (4.29) 式]. 因为当 $p \neq i$ 时, $\sum\limits_{j=1}^{b} n_{ij}n_{pj} = \lambda$ 以及 $n_{pj}^2 = n_{pj}$(因为 $n_{pj} = 0$ 或 1), 所以可将 (4.36) 式改写为

$$r(k-1)\hat{\tau}_i - \lambda\sum_{\substack{p=1 \\ p\neq 1}}^{a}\hat{\tau}_p = kQ_i \quad i = 1, 2, \cdots, a \tag{4.37}$$

最后, 注意到约束 $\sum\limits_{i=1}^{a}\hat{\tau}_i = 0$ 蕴含着 $\sum\limits_{\substack{p=1 \\ p\neq i}}^{a}\hat{\tau}_p = -\hat{\tau}_i$, 又因 $r(k-1) = \lambda(a-1)$, 故可得出

$$\lambda a\hat{\tau}_i = kQ_i \quad i = 1, 2, \cdots, a \tag{4.38}$$

因此, 平衡不完全区组模型的处理效应的最小二乘估计量是

$$\hat{\tau}_i = \frac{kQ_i}{\lambda a} \quad i = 1, 2, \cdots, a \tag{4.39}$$

例如, 考虑例 4.5 的平衡不完全区组设计. 因为 $Q_1 = -9/3$, $Q_2 = -7/3$, $Q_3 = -4/3$, $Q_4 = 20/3$, 得

$$\hat{\tau}_1 = \frac{3(-9/3)}{(2)(4)} = -9/8, \quad \hat{\tau}_2 = \frac{3(-7/3)}{(2)(4)} = -7/8$$

$$\hat{\tau}_3 = \frac{3(-4/3)}{(2)(4)} = -4/8, \quad \hat{\tau}_4 = \frac{3(20/3)}{(2)(4)} = 20/8$$

与 4.4.1 节所得的相同.

4.4.3 BIBD 中内部信息的恢复

在 4.4.1 节给出的平衡不完全区组设计的分析通常叫做**区组内分析** (intrablock analysis), 因为它消除了区组差异, 所以, 处理效应的全部对照都可以表示为同一区组内的观测值之间的比较. 这一分析对于固定区组以及随机区组都适用. Yates(1940) 提到, 当区组效应是均值为零、方差为 σ_β^2 的不相关随机变量时, 可以得出关于处理效应 τ_i 的附加信息. Yates 称获得这类附加信息的方法为**区组间分析** (interblock analysis).

将区组总和 $y_{.j}$ 看作为 b 个观测值的集合. 这些观测值的模型 [依照 John(1971)] 是

$$y_{.j} = k\mu + \sum_{i=1}^{a} n_{ij}\tau_i + \left(k\beta_j + \sum_{i=1}^{a} \epsilon_{ij}\right) \tag{4.40}$$

其中括号内的项可看作为误差项. μ 与 τ_i 的区组内估计量可以由最小二乘函数

$$L = \sum_{j=1}^{b} \left(y_{.j} - k\mu - \sum_{i=1}^{a} n_{ij}\tau_i\right)^2$$

最小化而得. 由此得到下面的最小二乘正规方程组:

$$\begin{aligned} &\mu : N\tilde{\mu} + r\sum_{i=1}^{a}\tilde{\tau}_i = y_{..} \\ &\tau_i : kr\tilde{\mu} + r\tilde{\tau}_i + \lambda\sum_{\substack{p=1\\ p\neq 1}}^{a}\tilde{\tau}_p = \sum_{j=1}^{b} n_{ij}y_{.j} \quad i = 1, 2, \cdots, a \end{aligned} \tag{4.41}$$

其中 $\tilde{\mu}$ 与 $\tilde{\tau}_i$ 表示**区组间估计量**. 加上约束 $\sum\limits_{i=1}^{a}\tilde{\tau}_i = 0$, 得方程组 (4.41) 的解为

$$\tilde{\mu} = \overline{y}_{..} \tag{4.42}$$

$$\tilde{\tau}_i = \frac{\sum\limits_{j=1}^{b} n_{ij}y_{.j} - kr\overline{y}_{..}}{r-\lambda} \quad i = 1, 2, \cdots, a \tag{4.43}$$

可以证明, 区组间估计量 $\{\tilde{\tau}_i\}$ 与区组内估计量 $\{\hat{\tau}_i\}$ 是不相关的.

区组间估计量 $\{\tilde{\tau}_i\}$ 可能不同于区组内估计量 $\{\hat{\tau}_i\}$. 例如, 算得例 4.5 的平衡不完全区组设计的区组间估计量为:

$$\tilde{\tau}_1 = \frac{663 - (3)(3)(72.50)}{3-2} = 10.50, \qquad \tilde{\tau}_2 = \frac{649 - (3)(3)(72.50)}{3-2} = -3.50$$

$$\tilde{\tau}_3 = \frac{652 - (3)(3)(72.50)}{3-2} = -0.50, \qquad \tilde{\tau}_4 = \frac{646 - (3)(3)(72.50)}{3-2} = -6.50$$

注意, 此处 $\sum\limits_{j=1}^{b} n_{ij}y_{.j}$ 的值前面在计算区组内分析中已调整的处理总和时用到过.

现在, 如果我们想要把区组间估计量和区组内估计量组合起来, 以期对于每一个 τ_i, 都可以得到单一、无偏、最小方差的估计. 可以证明, $\hat{\tau}_i$ 与 $\tilde{\tau}_i$ 都是无偏的, 而且

$$V(\hat{\tau}_i) = \frac{k(a-1)}{\lambda a^2}\sigma^2 \quad (\text{区组内})$$

$$V(\tilde{\tau}_i) = \frac{k(a-1)}{a(r-\lambda)}(\sigma^2 + k\sigma_\beta^2) \quad (\text{区组间})$$

我们取这两个估计量的一个线性组合, 比方说

$$\tau_i^* = \alpha_1\hat{\tau}_i + \alpha_2\tilde{\tau}_i \tag{4.44}$$

来估计 τ_i. 对这一种估计方法, 最小方差无偏组合估计量 τ_i^* 应有权重 $\alpha_1 = u_1/(u_1+u_2)$ 和 $\alpha_2 = u_2/(u_1+u_2)$, 其中 $u_1 = 1/\mathrm{V}(\hat{\tau}_i)$, $u_2 = 1/\mathrm{V}(\tilde{\tau}_i)$. 于是, 最优的权重与 $\hat{\tau}_i$ 与 $\tilde{\tau}_i$ 的方差成反比. 这意味着, 最好的组合估计量是

$$\tau_i^* = \frac{\hat{\tau}_i \frac{k(a-1)}{a(r-\lambda)}(\sigma^2 + k\sigma_\beta^2) + \tilde{\tau}_i \frac{k(a-1)}{\lambda a^2}\sigma^2}{\frac{k(a-1)}{\lambda a^2}\sigma^2 + \frac{k(a-1)}{a(r-\lambda)}(\sigma^2 + k\sigma_\beta^2)} \quad i = 1, 2, \cdots, a$$

可以把它简化为

$$\tau_i^* = \frac{kQ_i(\sigma^2 + k\sigma_\beta^2) + \left(\sum_{j=1}^{b} n_{ij}y_{.j} - kr\overline{y}_{..}\right)\sigma^2}{(r-\lambda)\sigma^2 + \lambda a(\sigma^2 + k\sigma_\beta^2)} \quad i = 1, 2, \cdots, a \tag{4.45}$$

然而, (4.45) 式不能用来估计 τ_i, 因为方差 σ^2 与 σ_β^2 是未知的. 通常的方法是利用测得的数据来估计 σ^2 与 σ_β^2, 然后用这些估计值代入 (4.45) 式中来代替这两个参数. σ^2 的估计量通常取区组内方差分析中的误差均方值, 即**区组内误差**. 于是有

$$\hat{\sigma}^2 = MS_E$$

σ_β^2 的估计值取自对处理已调整过的区组的均方值. 一般说来, 对于平衡不完全区组设计, 这一均方值是

$$MS_{\text{区组 (已调整的)}} = \frac{\left(\frac{k\sum_{i=1}^{a} Q_i^2}{\lambda a} + \sum_{j=1}^{b}\frac{y_{.j}^2}{k} - \sum_{i=1}^{a}\frac{y_{i.}^2}{r}\right)}{(b-1)} \tag{4.46}$$

它的期望值 [由 Graybill(1961) 导出] 是

$$\mathrm{E}[MS_{\text{区组 (已调整的)}}] = \sigma^2 + \frac{a(r-1)}{b-1}\sigma_\beta^2$$

这样一来, 当 $MS_{\text{区组 (已调整的)}} > MS_E$ 时, $\hat{\sigma}_\beta^2$ 的估计量是

$$\hat{\sigma}_\beta^2 = \frac{[MS_{\text{区组 (已调整的)}} - MS_E](b-1)}{a(r-1)} \tag{4.47}$$

且当 $MS_{\text{区组 (已调整的)}} \leqslant MS_E$ 时, 令 $\hat{\sigma}_\beta^2$=0. 这就得出组合估计

$$\tau_i^* = \begin{cases} \dfrac{kQ_i(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2) + \left(\sum_{j=1}^{b} n_{ij}y_{.j} - kr\overline{y}_{..}\right)\hat{\sigma}^2}{(r-\lambda)\hat{\sigma}^2 + \lambda a(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2)}, & \hat{\sigma}_\beta^2 > 0 \quad (4.48\text{a}) \\ \dfrac{y_{i.} - (1/a)y_{..}}{r}, & \hat{\sigma}_\beta^2 = 0 \quad (4.48\text{b}) \end{cases}$$

今对例 4.5 的数据计算这些组合估计量. 由表 4.24 得 $\hat{\sigma}^2 = MS_E = 0.65$ 以及 MS=22.03(在计算 $MS_{\text{区组 (已调整的)}}$ 时, 我们利用了对称设计. 一般说来, 须用 (4.46) 式). 因为 $MS_{\text{区组 (已调整的)}} \geqslant MS_E$, 用 (4.47) 式估计 $\hat{\sigma}_\beta^2$ 得

$$\hat{\sigma}_\beta^2 = \frac{(22.03 - 0.65)(3)}{4(3-1)} = 8.02$$

因此, 可将 $\hat{\sigma}^2 = 0.65$ 和 $\hat{\sigma}_\beta^2 = 8.02$ 代入 (4.48a) 式得出组合估计量, 如下表所示. 为方便起见, 亦列出区组内估计量和区组间估计量. 该例中, 组合估计接近区组内估计, 因为区组间估计的方差相对较大.

参数	区组内估计值	区组间估计值	组合估计
τ_1	−1.12	10.50	−1.09
τ_2	−0.88	−3.50	−0.88
τ_3	−0.50	−0.50	−0.50
τ_4	2.50	−6.50	2.47

4.5 思 考 题

4.1 一位化学家想要检验 4 种化学试剂对于一种特定类型布料的强度效应. 因为布匹之间会有变异性, 这位化学家决定用随机化区组设计, 将布匹考虑为区组. 她选取了 5 匹并用所有 4 种化学制剂随机地对每匹进行试验. 测得的抗拉强度如下. 分析这些数据 (用 $\alpha = 0.05$) 并写出恰当的结论.

化学试剂	布匹				
	1	2	3	4	5
1	73	68	74	71	67
2	73	67	75	72	70
3	75	68	78	73	68
4	73	71	75	75	69

*4.2 研究 3 种不同的洗涤液对于 5 加仑的牛奶容器内细菌增长的抑制作用. 分析是在实验室内做的, 每天只能做 3 个试验, 因为时间可能表示一种潜在的变异性来源, 实验者决定用随机化区组设计. 4 天所得数据如下. 分析该实验的数据 (用 $\alpha = 0.05$) 并写出结论.

洗涤液	时间 (天)			
	1	2	3	4
1	13	22	18	39
2	16	24	17	44
3	5	4	1	22

4.3 对思考题 4.1 中每种化学试剂的平均抗拉强度绘图, 并与适当尺度的 t 分布进行比较. 从图中你能得出什么结论?

4.4 对思考题 4.2 中每种洗涤液的平均细菌量绘图, 并与适当尺度的 t 分布进行比较. 你得出什么结论?

*4.5 考虑 4.1 节中描述的硬度检验实验. 假定实验按照描述的进行, 以下是洛氏硬度 C 标尺下的数据 (减去 40 个单位进行规范):

压头	试件			
	1	2	3	4
1	9.3	9.4	9.6	10.0
2	9.4	9.3	9.8	9.9
3	9.2	9.4	9.5	9.7
4	9.7	9.6	10.0	10.2

(a) 分析该实验中的数据.

(b) 用 Fisher LSD 方法作 4 个压头间的比较, 确认哪个压头的硬度读数均值有显著不同.

(c) 分析该实验的残差.

*4.6 消费品公司准备把直接邮寄宣传材料作为广告宣传的主要手段. 宣传材料中提供了 3 种不同的设计方案, 公司想估计它们的效用, 因为 3 种设计的成本可能存在显著差异. 公司准备在国内的 4 个不同区域邮寄 5 000 个样本给每个潜在顾客来检验这 3 种设计. 因为已知每个区域的顾客基数, 所以考虑区域为区组. 每次邮寄的响应数如下:

设计	区域			
	东北	西北	东南	西南
1	250	350	219	375
2	400	525	390	580
3	275	340	200	310

(a) 分析该实验中的数据.

(b) 用 Fisher LSD 方法作 3 个设计间的比较, 确认哪个设计的平均响应率有显著不同.

(c) 分析该实验的残差.

4.7 研究 3 种润滑油对柴油发动机货车节省燃油的效应. 用发动机运行 15 分钟后的制动油耗率来度量燃油的经济性. 研究中用了 5 台不同的发动机, 实验者利用了完全随机化区组设计.

油	发动机				
	1	2	3	4	5
1	0.500	0.634	0.487	0.329	0.512
2	0.535	0.675	0.520	0.435	0.540
3	0.513	0.595	0.488	0.400	0.510

(a) 分析该实验中的数据.

(b) 用 Fisher LSD 方法作 3 种润滑油间的比较, 确认哪种润滑油的制动油耗率有显著不同.

(c) 分析该实验的残差.

4.8 在*Fire Safety Journal* 上的一篇论文 ("The Effect of Nozzle Design on the Stability and Performance of Turbulent Water Jets," Vol. 4, August 1981) 描述了一个实验, 其中的形状因子是由几种不同的喷嘴设计决定的, 每种喷嘴设计考虑了喷流速度的 6 个水平. 问题的焦点是将喷流速度看作讨厌变量时, 喷嘴设计间有无潜在的差异. 数据如下:

喷嘴设计	喷流速度 (m/s)					
	11.73	14.37	16.59	20.43	23.46	28.74
1	0.78	0.80	0.81	0.75	0.77	0.78
2	0.85	0.85	0.92	0.86	0.81	0.83
3	0.93	0.92	0.95	0.89	0.89	0.83
4	1.14	0.97	0.98	0.88	0.86	0.83
5	0.97	0.86	0.78	0.76	0.76	0.75

(a) 喷嘴设计影响形状因子吗? 用盒图与方差分析来比较喷嘴设计, 用 $\alpha = 0.05$.

(b) 分析这一实验的残差.

(c) 对于形状因子, 哪些喷嘴设计有差异? 描绘每种喷嘴设计的平均形状因子并与适当尺度的 t 分布进行比较. 将你从这一图中所得的结论与从 Duncan 多重极差检验法中所得的结论进行比较.

4.9 考虑 3.8 节描述的比值控制法实验. 这一实验实际上是用随机化区组设计进行的, 其中选取 6 个时间周期作为区组, 在每一时间周期中检验所有 4 种比值控制法. 平均电解槽电压与每一电解槽电压的标准差 (圆括号内的数) 如下:

比值控制算法	时间周期					
	1	2	3	4	5	6
1	4.93 (0.05)	4.86 (0.04)	4.75 (0.05)	4.95 (0.06)	4.79 (0.03)	4.88 (0.05)
2	4.85 (0.04)	4.91 (0.02)	4.79 (0.03)	4.85 (0.05)	4.75 (0.03)	4.85 (0.02)
3	4.83 (0.09)	4.88 (0.13)	4.90 (0.11)	4.75 (0.15)	4.82 (0.08)	4.90 (0.12)
4	4.89 (0.03)	4.77 (0.04)	4.94 (0.05)	4.86 (0.05)	4.79(0.03)	4.76 (0.02)

(a) 分析平均电解槽电压数据 (用 $\alpha = 0.05$). 比值控制法的选择影响平均电解槽电压吗?

(b) 用适当的分析法来分析电压的标准差 (称之为 "电位噪音"), 比值控制法的选择影响电位噪音吗?

(c) 用适当的残差分析法进行分析.

(d) 如果想要同时减少平均电解槽电压与电位噪音, 你将选择哪种比值控制法?

*4.10 一家铝合金厂生产锭状的晶粒细化剂. 工厂用 4 种炉子生产产品. 每种炉子都有它独特的抽检特性, 所以, 在有多于一种炉子的铸造车间中进行任一实验就要将炉子看作讨厌变量. 过程工程师预料搅拌率会影响产品晶粒的大小. 每种炉子以 4 种不同的搅拌率运行. 对一种指定的细化剂用随机化区组设计进行试验, 所得的晶粒大小数据如下所示.

搅拌率 (转/分)	炉子			
	1	2	3	4
5	8	4	5	6
10	14	5	6	9
15	14	6	9	2
20	17	9	3	6

(a) 是否有证据说明搅拌率对晶粒大小有影响?

(b) 在正态概率纸上画出这一实验的残差图. 解释该图.

(c) 画出残差与炉子以及残差与搅拌率的关系图. 该图能提供有用的信息吗?

(d) 如果要求细小的晶粒大小, 对这种指定的晶粒细化剂, 过程工程师应该推荐选用哪种搅拌率和炉子呢?

*4.11 用一般回归的显著性检验法分析思考题 4.2 中的数据.

*4.12 假定化学试剂类型和布匹是固定的, 估计思考题 4.1 中的模型参数 τ_i 与 β_j.

4.13 画出思考题 4.2 的抽检特性曲线. 这一检验法对处理效应的细小差异的灵敏度如何?

4.14 设思考题 4.1 中化学试剂 2 与布匹 3 的观测值缺失. 用缺失值的估计值来分析这一问题. 再用精确的分析法并比较这些结果.

4.15 考虑思考题 4.5 中的硬度检验实验. 假设试件 3 与压头 2 的观测值缺失. 用缺失值的估计值来分析这一问题.

4.16 **随机化区组中的两个缺失值**. 假定在思考题 4.1 中化学制剂 2 与布匹 3 以及化学制剂 4 与布匹 4 的观测值都缺失.

(a) 用 4.1.3 节所述的迭代估计法估计缺失值, 并分析这一设计.

(b) 将 SS_E 对那两个缺失值求导并使之等于零, 从而解得缺失值的估计值. 用这两个缺失值的估计值来分析这一设计.

(c) 当观测值在不同的区组中时, 导出估计两个缺失值的一般公式.

(d) 当观测值在相同的区组中时, 导出估计两个缺失值的一般公式.

4.17 工业工程师进行关于眼睛聚焦时间的实验. 他感兴趣于目标与眼睛的距离关于聚焦时间的效应. 研究 4 种不同的距离. 他用 5 个对象来做实验. 因为各个对象间会有差异, 他决定用随机化区组设计来做实验. 所得数据如下. 分析该实验数据 (用 $\alpha = 0.05$) 并写出恰当的结论.

距离 (英尺)	对象				
	1	2	3	4	5
4	10	6	6	6	6
6	7	6	6	1	6
8	5	3	3	2	5
10	6	4	4	2	3

4.18 研究 5 种不同的配料 (A, B, C, D, E) 对某一化学过程反应时间的效应. 每批新材料仅够进行 5 次试验. 每次试验大约需要 $1\frac{1}{2}$ 小时, 所以一天只能做 5 次试验. 实验者决定用拉丁方来进行实验, 使日期和批次效应可以系统地控制. 得出数据如下. 分析该实验的数据 (用 $\alpha = 0.05$) 并写出结论.

批次	时间				
	1	2	3	4	5
1	$A=8$	$B=7$	$D=1$	$C=7$	$E=3$
2	$C=11$	$E=2$	$A=7$	$D=3$	$B=8$
3	$B=4$	$A=9$	$C=10$	$E=1$	$D=5$
4	$D=6$	$C=8$	$E=6$	$B=6$	$A=10$
5	$E=4$	$D=2$	$B=3$	$A=8$	$C=8$

*4.19 一位工业工程师研究 4 种组装方法 (A, B, C, D) 对彩色电视组件组装时间的效应. 选取 4 位操作工来做. 工程师还知道, 每种组装方法都产生这样的疲劳, 即最后一道组装所需要的时间大于第一道组装所需要的时间, 无论哪种方法均是如此. 也就是说, 所需的组装时间有趋向性. 考虑到这个变异性来源, 工程师用如下面所示的拉丁方设计. 分析该实验数据 (用 $\alpha = 0.05$) 并写出恰当的结论.

组装次序	操作工			
	1	2	3	4
1	$C=10$	$D=14$	$A=7$	$B=8$
2	$B=7$	$C=18$	$D=11$	$A=8$
3	$A=5$	$B=10$	$C=11$	$D=9$
4	$D=10$	$A=10$	$B=12$	$C=14$

4.20 设思考题 4.18 中的批次 3 和时间 4 的观测值缺失. 用 (4.24) 式估计这一缺失值, 并用此值进行分析.

4.21 考虑具有固定效应的行 (α_i)、列 (β_k) 和处理 (τ_j) 的 $p \times p$ 拉丁方. 给出模型参数 α_i, β_k, τ_j 的最小二乘估计.

4.22 导出拉丁方设计的缺失值公式 [(4.24) 式].

4.23 **涉及多个拉丁方的设计**. [见 Cochran and Cox(1957), John(1971)]$p \times p$ 拉丁方对每一处理仅含有 p 个观测值. 为了得到更多的重复, 实验者可以用多个 (比方说 n 个) 拉丁方. 用相同的或不同的拉丁方是无关重要的. 合适的模型是

$$y_{ijkh} = \mu + \rho_h + \alpha_{i(h)} + \tau_j + \beta_{k(h)} + (\tau\rho)_{jh} + \varepsilon_{ijkh} \quad \begin{cases} i = 1, 2, \cdots, p \\ j = 1, 2, \cdots, p \\ k = 1, 2, \cdots, p \\ h = 1, 2, \cdots, n \end{cases}$$

其中 y_{ijkh} 是处理 j 在第 h 个拉丁方的行 i 和列 k 上的观测值. $\alpha_{i(h)}$ 和 $\beta_{k(h)}$ 是第 h 个拉丁方的行效应和列效应, ρ_h 是第 h 个拉丁方的效应, $(\tau\rho)_{jh}$ 是处理和拉丁方之间的交互作用.

(a) 建立这一模型的正规方程组, 并解出模型参数的估计值. 假定关于这些参数的恰当的边界条件是, 对每一 h 有 $\sum_h \hat{\rho}_h = 0$, $\sum_i \hat{\alpha}_{i(h)} = 0$ 以及 $\sum_k \beta_{k(h)} = 0$; 对每一 h 有 $\sum_j \hat{\tau}_j = 0$, $\sum_j (\hat{\tau}\rho)_{jh} = 0$; 对每一 j 有 $\sum_h (\hat{\tau}\rho)_{jh} = 0$.

(b) 写出这一设计的方差分析表.

4.24 讨论怎样把附录中的抽检特性曲线用到拉丁方设计中去.

4.25 假定在思考题 4.18 中, 时间 5 所取的数据得出了不正确的分析并不得不放弃这些数据. 研究其余数据的恰当的分析法.

4.26 测量一个化学过程的产量, 其中用 5 批原材料、5 种酸性浓度、5 种持续时间 (A, B, C, D, E), 以及 5 种催化剂浓度 $(\alpha, \beta, \gamma, \delta, \varepsilon)$. 用下面的正交拉丁方. 分析该实验数据 (用 $\alpha = 0.05$) 并写出结论.

批　次	酸性浓度				
	1	2	3	4	5
1	$A\alpha$=26	$B\beta$=16	$C\gamma$=19	$D\delta$=16	$E\varepsilon$=13
2	$B\gamma$=18	$C\delta$=21	$D\varepsilon$=18	$E\alpha$=11	$A\beta$=21
3	$C\varepsilon$=20	$D\alpha$=12	$E\beta$=16	$A\gamma$=25	$B\delta$=13
4	$D\beta$=15	$E\gamma$=15	$A\delta$=22	$B\varepsilon$ =14	$C\alpha$=17
5	$E\delta$=10	$A\varepsilon$=24	$B\alpha$=17	$C\beta$=17	$D\gamma$=14

4.27 假设在思考题 4.19 中, 工程师推测, 4 位操作工所在的工作地点亦可能是另一个变异性来源. 为此, 可以引入第 4 个因子 [即工作地点 $(\alpha, \beta, \gamma, \delta)$], 并进行另一试验, 得到如下的正交拉丁方. 分析这些数据 (用 $\alpha = 0.05$) 并写出结论.

组装次序	操　作　工			
	1	2	3	4
1	$C\beta$=11	$B\gamma$=10	$D\delta$=14	$A\alpha$=8
2	$B\alpha$=8	$C\delta$=12	$A\gamma$=10	$D\beta$=12
3	$A\delta$=9	$D\alpha$=11	$B\beta$=7	$C\gamma$=15
4	$D\gamma$=9	$A\beta$=8	$C\alpha$=18	$B\delta$=6

4.28 构造一个可用来研究 5 个因子的效应的 5×5 超方. 给出这一设计的方差分析表.

4.29 考虑思考题 4.19 与思考题 4.27 中的数据. 删去思考题 4.27 中的希腊字母, 用思考题 4.23 中所研究的方法分析这些数据.

4.30 考虑思考题 4.15 中带有一个缺失值的随机化区组设计. 用 4.1.4 节中讨论的缺失值问题的精确分析法来分析这些数据. 将你的结果与思考题 4.15 给出的对这些数据的近似分析法的结果进行比较.

*4.31 一位工程师研究 5 种类型的汽油添加剂的性能特征. 在公路实验中他用汽车作为区组. 因为有时间限制, 他必须用不完全区组设计. 他用 5 个区组进行了以下的平衡区组设计. 分析该实验数据 (用 $\alpha = 0.05$) 并写出结论.

添加剂	汽　　车				
	1	2	3	4	5
1		17	14	13	12
2	14	14		13	10
3	12		13	12	9
4	13	11	11	12	
5	11	12	10		8

*4.32 对思考题 4.31 中的数据构造正交对照. 计算每一对照的平方和.

*4.33 研究 7 种不同的硬木浓度, 以便确定它们对所生产的纸张强度的效应. 但是, 实验装置每天只能进行 3 次试验, 因为实验日期可能不同, 化验员用平衡不完全区组设计如下. 分析该实验数据 (用 $\alpha = 0.05$) 并写出结论.

硬木浓度 (%)	日期						
	1	2	3	4	5	6	7
2	114				120		117
4	126	120				119	
6		137	117				134
8	141		129	149			
10		145			150	143	
12			120		118	123	
14				136		130	127

*4.34 用一般回归的显著性检验法分析例 4.5 的数据.

*4.35 证明 $k\sum_{i=1}^{a} Q_i^2/(\lambda a)$ 是 BIBD 中已调整的处理平方和.

*4.36 实验者想用有两次试验的区组来比较 4 种处理. 给这一实验找一个有 6 个区组的平衡不完全区组设计.

4.37 实验者想用有 4 次试验的区组来比较 8 种处理. 找出有 14 个区组和 $\lambda = 3$ 的平衡不完全区组设计.

4.38 对思考题 4.31 的设计进行区组间分析.

4.39 对思考题 4.33 的设计进行区组间分析.

*4.40 证明参数为 $a = 8$, $r = 8$, $k = 4$, $b = 16$ 的平衡不完全区组设计不存在.

*4.41 证明区组内估计量 $\{\hat{\tau}_i\}$ 的方差是 $k(a-1)\sigma^2/(\lambda a^2)$.

4.42 **扩展的不完全区组设计.** 有时, 区组的大小服从关系式 $a < k < 2a$. 扩展的不完全区组设计由下列项组成: 每个区组中的每个处理的一次重复, 以及 $k^* = k - a$ 的一个不完全区组设计组成. 在平衡情形下, 不完全区组设计将有参数 $k^* = k - a$, $r^* = r - b$ 以及 λ^*. 写出这个统计分析. (提示: 在扩展的不完全区组设计中, 有 $\lambda = 2r - b + \lambda^*$.)

第 5 章　析因设计导引

本章纲要

5.1　基本定义与原理
5.2　析因设计的优点
5.3　二因子析因设计
5.3.1　一个例子
5.3.2　固定效应模型的统计分析
5.3.3　模型适合性检验
5.3.4　估计模型参数
5.3.5　样本量的选择
5.3.6　假定在二因子模型中没有交互作用
5.3.7　每单元一个观测值
5.4　一般的析因设计
5.5　拟合响应曲线与曲面
5.6　析因设计中的区组化
第 5 章补充材料
S5.1　二因子析因设计中的期望均方
S5.2　交互作用的定义
S5.3　二因子析因设计模型中的可估函数
S5.4　二因子析因设计中的回归模型方法
S5.5　模型的分层

5.1　基本定义与原理

很多实验要研究两个或多个因子的效应. 一般地, **析因设计**(factorial design) 对这类实验是最有效的. 所谓析因设计是指, 在这类实验的每一次完全试验或每一次重复中, 这些因子水平的所有可能的组合都被研究到. 例如, 当因子 A 有 a 个水平和因子 B 有 b 个水平时, 则每次重复都包含全体 ab 个处理组合. 当这些因子被安排在某一析因设计中时, 常称为是**交叉的**.

一个因子的效应定义为当这一因子的水平改变时所产生响应的变化. 这种效应常称为**主效应**(main effect), 因为它来自实验中所感兴趣的基本因子. 例如, 考虑图 5.1 的简单实验. 这是二因子二水平的析因实验. 我们称这些水平为 "低" 和 "高", 分别记为 "−" 和 "+". 该二水平设计中, 因子 A 的主效应可看作为 A 的低水平的平均响应和 A 的高水平的平均响应之间的差. 数值是

$$A=\frac{40+52}{2}-\frac{20+30}{2}=21$$

这就是说, 因子 A 从低水平增至高水平使得其平均响应增加了 21 个单位. 类似地, B 的主效应是

$$B=\frac{30+52}{2}-\frac{20+40}{2}=11$$

当这些因子多于两个水平时, 必须修改上述方法, 因为有其他方式来定义因子的效应. 这一点随后进行更全面的讨论.

在一些实验中, 我们会发现, 一个因子的水平间的响应差随其他因子的水平不同而不同. 当这种情况出现时, 因子之间就存在**交互作用**(interaction). 例如, 考虑图 5.2 的数据. 对因子 B 的低水平 (即 B^-), A 的效应是

$$A=50-20=30$$

而对 B 的高水平 (即 B^+), A 的效应是

$$A = 12 - 40 = -28$$

因为 A 的效应依赖于因子 B 所选的水平, 可见 A 与 B 之间存在交互作用. 交互作用的大小是这两个 A 效应的平均差, 即 $AB = (-28 - 30)/2 = -29$. 显然, 在该实验中这个交互作用是比较大的.

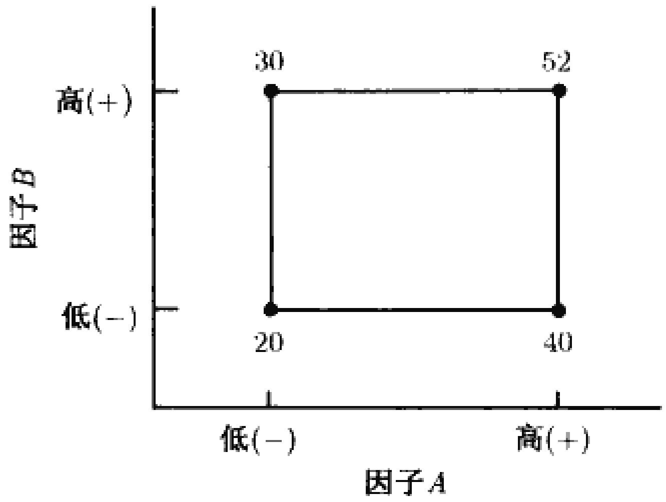


图 5.1 二因子析因实验, 其响应 (y) 显示在各角点上

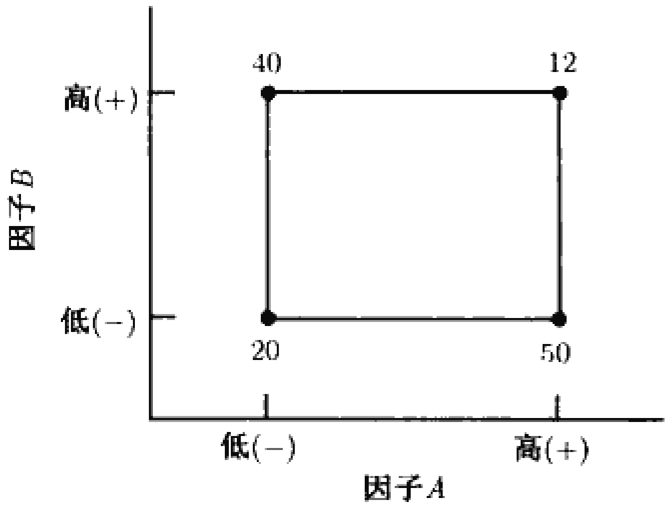


图 5.2 有交互作用的二因子析因实验

这些想法可用图解来说明. 图 5.3 是图 5.1 的响应数据图, 是因子 A 对因子 B 的各水平的. B^- 直线与 B^+ 直线几乎平行, 表明因子 A 与 B 之间没有交互作用. 相似地, 图 5.4 是图 5.2 的响应数据图. 由图看出, B^- 直线与 B^+ 直线不平行, 表明因子 A 与 B 之间有交互作用. 这样的图形在解释显著性交互作用和报告结果给未受过专门统计培训的管理者方面常常很有用. 不过, 这不可以用作为数据分析的唯一方法, 因为它的解释是主观的, 它的表面现象常常会令人误解.

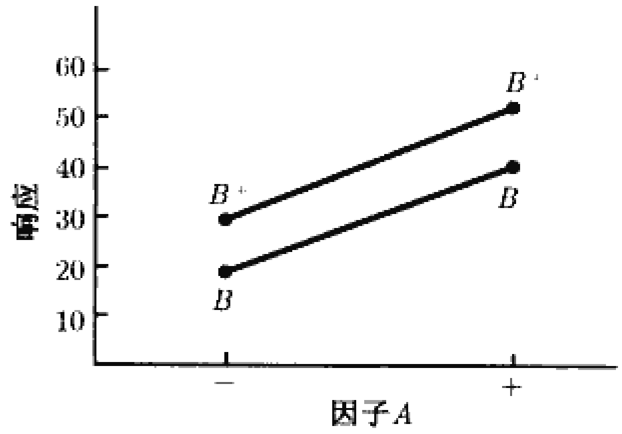


图 5.3 无交互作用的析因实验

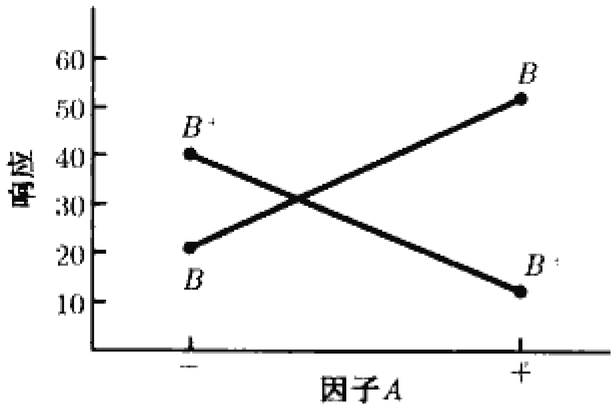


图 5.4 有交互作用的析因实验

也可用另一种方法来说明交互作用的概念. 假定两个设计因子都是**定量的** (例如, 温度、气压、时间等). 此时, 二因子析因设计的**回归模型表达式**可以写成

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

其中, y 是响应, β 是待定参数, 变量 x_1 表示因子 A, 变量 x_2 表示因子 B, ϵ 是随机误差项. 变量 x_1 和 x_2 用 -1 到 $+1$ 的**规范化**定义 (A 与 B 的低水平和高水平), x_1x_2 表示 x_1 和 x_2 间的交互作用.

回归模型的参数估计值可以根据相应的效应估计值进行计算. 图 5.1 的实验中, 我们发现 A 与 B 的主效应分别为 A= 21, B=11, β_1 和 β_2 的估计值是相应的主效应的值的一半, 因此, $\hat{\beta}_1 = 21/2 = 10.5$, $\hat{\beta}_2 = 21/2 = 5.5$. 图 5.1 的交互作用效应是 $AB = 1$, 所以, 回归模型中的交互作用的系数的值为 $\hat{\beta}_{12} = 1/2 = 0.5$. 参数 β_0 用所有 4 个响应的均值来估计, 即 $\hat{\beta}_0 = (20 + 40 + 30 + 52)/4 = 35.5$. 因此, 拟合回归模型为

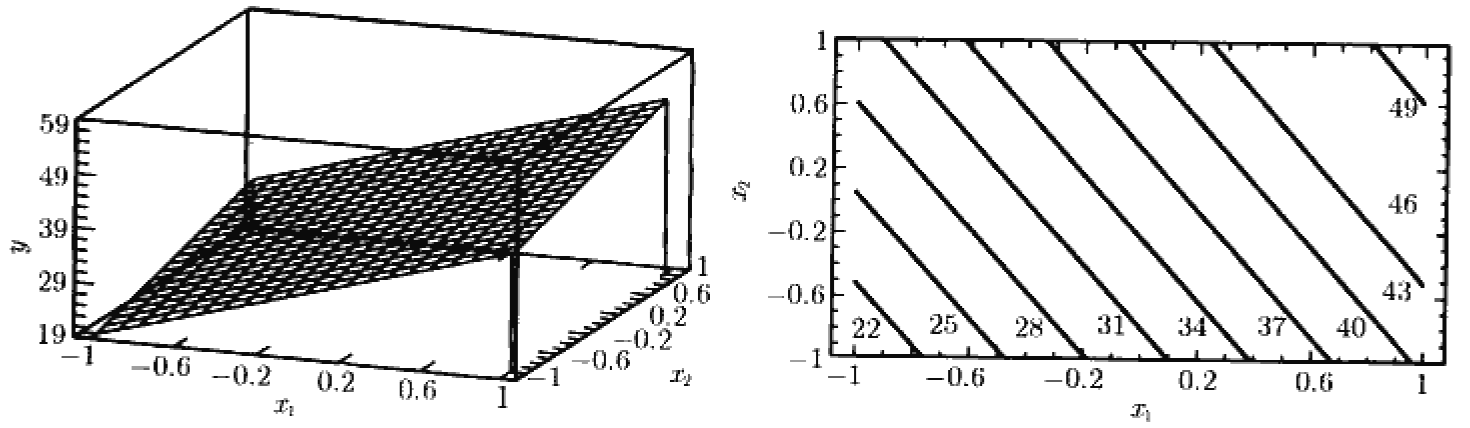
$$\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 0.5x_1x_2$$

从两水平 (− 和 +) 析因设计得到的参数估计实际上就是**最小二乘估计**(更多内容见本章后面).

交互作用系数 ($\hat{\beta}_{12} = 0.5$) 相对于主效应系数 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 而言是比较小的. 我们将此看作是交互作用小得可以忽略不计. 因此, 去掉项 $0.5x_1x_2$, 模型变为

$$\hat{y} = 35.5 + 10.5x_1 + 5.5x_2$$

图 5.5 给出了该模型的图形表示. 图 5.5a 给出了由 x_1 和 x_2 的不同组合生成的 y 值的平面图. 这个 3 维图称为**响应曲面图**(response surface plot). 图 5.5b 显示了 x_1x_2 平面中的连续响应 y 的等高线图. 因为响应曲面是一个平面, 等高线都是平行直线.



(a) 响应曲面　　(b) 等高线图

图 5.5　模型 $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2$ 的响应曲面和等高线图

假定交互作用对实验的贡献是不能忽略的, 也就是, 系数 β_{12} 并不小. 图 5.6 给出了模型

$$\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 8x_1x_2$$

的响应曲面以及等高线图 (我们令交互作用效应为两主效应的均值). 注意到显著的交互作用效应 "扭曲" 了图 5.6a 中的平面. 图 5.6b 显示了由平面的扭曲导致的 x_1x_2 平面的连续响应的弯曲等高线. **交互作用**在该实验的潜在响应曲面中表现为**弯曲的形式**.

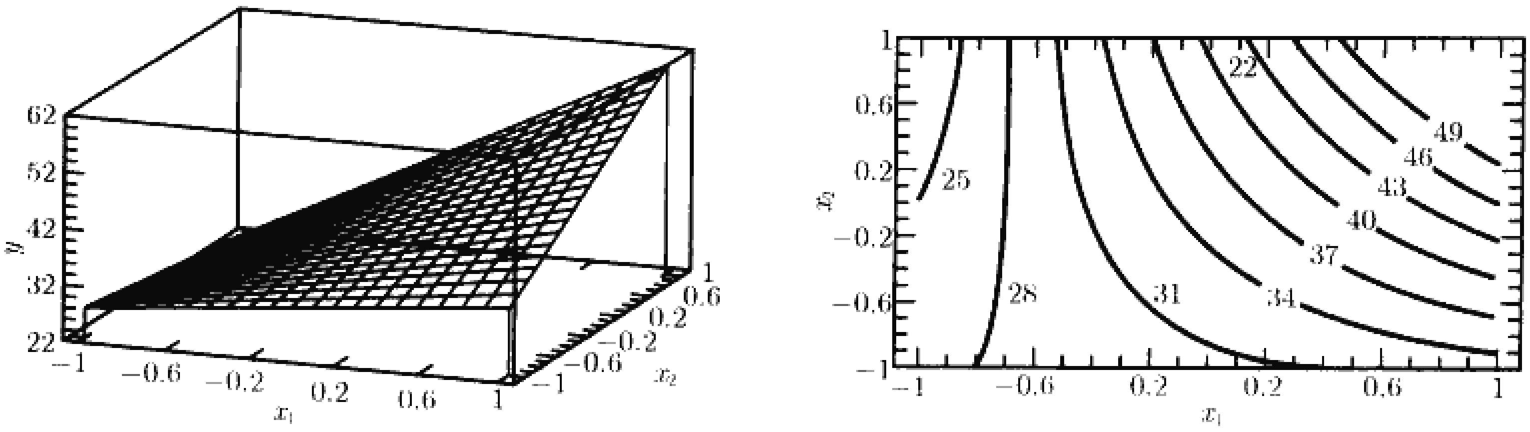
实验的响应曲面模型是极其重要的和极其有用的. 关于这一点, 我们将在 5.5 节和随后的章节中详细叙述.

一般地, 当交互作用较大时, 对应的主效应的实际作用较小. 对图 5.2 的实验, A 的主效应的估计是

$$A = \frac{50 + 12}{2} - \frac{20 + 40}{2} = 1$$

这很小, 我们试图作出结论, 因子 A 不起作用. 然而, 当我们对因子 B 的不同水平来检查 A 的效应时, 情况并非如此. 因子 A 起作用, 但是它依赖于因子 B 的水平. 也就是说, AB 交互作用

的信息比主效应的更为有用. 显著的交互作用经常会掩盖主效应的显著性. 图 5.4 中的交互作用图清楚地指明了这一点. 在出现显著的交互作用时, 实验者通常必须固定其他因子的各个水平来检查一个因子 (比方说 A), 以便得出关于 A 的主效应的结论.



(a) 响应曲面　　(b) 等高线图

图 5.6 模型 $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 8x_1x_2$ 的响应曲面和等高线图

5.2 析因设计的优点

容易说明析因设计的优点. 设有两个因子 A 与 B, 它们各有两个水平. 因子的水平记为 A^-, A^+, B^- 与 B^+. 关于两个因子的信息可以用一次变化一个因子的方法来取得. 如图 5.7 所示. 变动因子 A 的效应由 $A^+B^- - A^-B^-$ 给出, 变动因子 B 的效应由 $A^-B^+ - A^-B^-$ 给出. 因为有实验误差, 所以在每一处理组合处取两个观测值并用平均响应来估计因子的效应. 这样一来, 总共需要 6 个观测值.

如果实施一个析因实验, 当然还要增加一个处理组合 A^+B^+. 现在, 只用 4 个观测值, 就可得到 A 效应的两个估计: $A^+B^- - A^-B^-$ 与 $A^+B^+ - A^-B^+$. 同样, 亦得出 B 效应的两个估计. 将每一主效应的两个估计取其平均值即得平均主效应, 其结果的**精确度**和由单因子实验所得的相同, 但只用了 4 个观测值. 从而, 我们可以说, 析因设计相对于一次一因子实验的效率是 $6/4 = 1.5$. 一般说来, 当因子的个数增加时, 相对效率亦增加, 如图 5.8 所示.

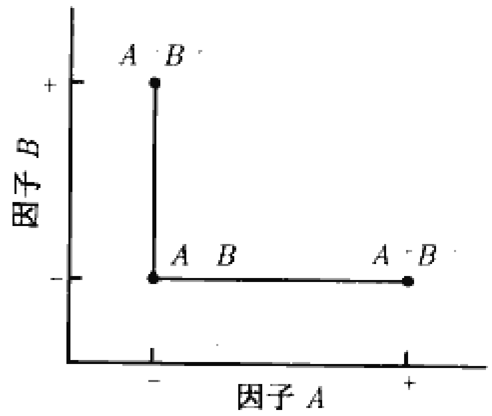


图 5.7 一次一因子实验

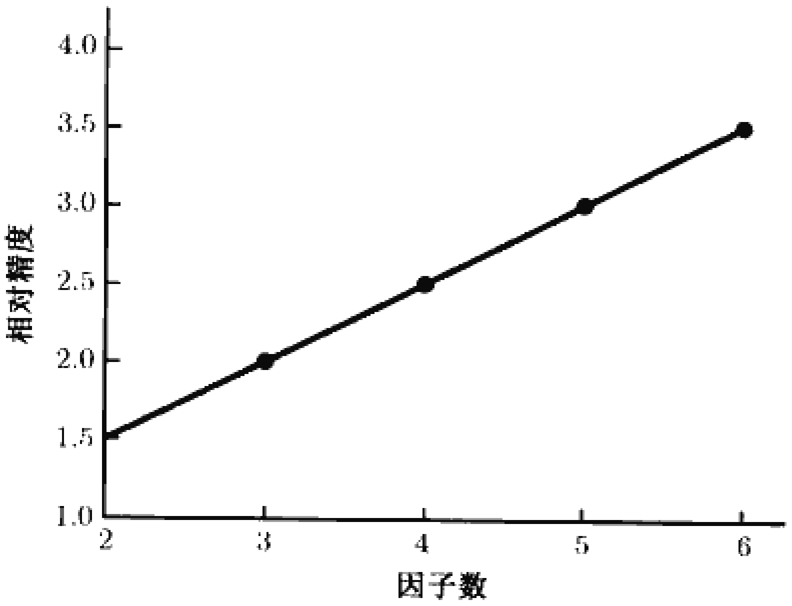


图 5.8 析因设计对一次一因子实验的相对精度 (二因子水平)

今假定有交互作用. 如果一次一因子设计表明 A^-B^+ 与 A^+B^- 给出比 A^-B^- 较好的响应的话, 则逻辑推论是 A^+B^+ 甚至会更好. 然而, 当有交互作用时, 这一结论可能有*严重错误*. 例如, 请参阅图 5.2 的实验.

简要说来, 析因设计有几个优点, 它比一次一因子实验效率高. 当有交互作用时, 为避免产生令人误解的结论, 必须用析因设计. 最后, 析因设计容许一个因子相对于其他各因子的几个水平来估计其效应, 所得结论在实验条件的范围内是有效的.

5.3　二因子析因设计

5.3.1　一个例子

最简单的析因设计只含有两个因子或两个处理组. 因子 A 有 a 个水平, 因子 B 有 b 个水平, 将它们安排到析因设计中. 也就是说, 实验的每次重复都含有 ab 个处理组合. 一般说来, 共有 n 次重复.

举一个含有两个因子的析因设计的例子, 一位工程师设计一种用在某装置内的电池, 该装置将遭受温度忽高忽低的极端变化. 他这时能够选择的唯一设计参数就是电池的板极材料, 有 3 种可能的选择. 当安装好这种装置并送到现场去之后, 这位工程师就再也不能控制这一装置所遇到的温度的极端变化了, 经验告诉他, 温度有可能影响电池的有效寿命. 但是, 为了试验目的所需, 在生产研制实验中, 温度是可以控制的.

工程师决定在 3 个温度水平 (15℉, 70℉和 125℉) 上检验所有 3 种板极材料, 这 3 种温度与产品使用的环境温度相符. 因为有两个因子, 每个因子有 3 个水平, 所以该设计有时称为 **3^2析因设计**. 在每种板极材料与温度组合上检验 4 节电池, 依随机次序进行所有 36 次检验. 实验和观测得到的电池寿命数据如表 5.1 所示.

表 5.1　电池设计例子的寿命 (单位: 小时) 数据

材料类型	温度 (℉)					
	15		70		125	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

在这一问题中, 工程师想要回答下述问题:

(1) 材料类型和温度对电池的寿命有何影响?

(2) 是否能选出一种材料使得*不论温度高低都能使电池有长的寿命*?

后一问题特别重要. 有可能寻求一种受温度影响较小的材料. 如果能做到这一点, 工程师就能生产出一种能经受外界温度变化的耐用电池. 这是一个为**稳健产品设计**(robust product design)而使用统计实验设计的例子, 而稳健产品设计是一个十分重要的工程问题.

这一设计是一般二因子析因设计的特例. 一般说来, 令 y_{ijk} 表示因子 A 取第 i 水平 $(i=1,2,\ldots,a)$、因子 B 取第 j 水平 $(j=1,2,\ldots,b)$ 时第 k 次重复 $(k=1,2,\ldots,n)$ 的响应观测值. 一般情况出现的两因子析因实验如表 5.2 所示. abn 个观测值的顺序是随机选择的, 所以, 这一设计是一个**完全随机化设计**.

表 5.2 二因子设计的一般排列表

		因子 B			
		1	2	$\cdots$	b
因子 A	1	$y_{111},y_{112},\cdots,y_{11n}$	$y_{121},y_{122},\cdots,y_{12n}$		$y_{1b1},y_{1b2},\cdots,y_{1bn}$
	2	$y_{211},y_{212},\cdots,y_{21n}$	$y_{221},y_{222},\cdots,y_{22n}$		$y_{2b1},y_{2b2},\cdots,y_{2bn}$
	$\vdots$				
	a	$y_{a11},y_{a12},\cdots,y_{a1n}$	$y_{a21},y_{a22},\cdots,y_{a2n}$		$y_{ab1},y_{ab2},\cdots,y_{abn}$

这些析因设计的观测值可以用一个模型来描述. 可以用多种方法写出析因实验的模型, **效应模型**为

$$y_{ijk}=\mu+\tau_i+\beta_j+(\tau\beta)_{ij}+\varepsilon_{ijk}\quad\begin{cases}i=1,2,\cdots,a\\ j=1,2,\cdots,b\\ k=1,2,\cdots,n\end{cases}\tag{5.1}$$

其中 μ 是总平均效应, τ_i 是行因子 A 的第 i 水平的效应, β_j 是列因子 B 的第 j 水平的效应, $(\tau\beta)_{ij}$ 是 τ_i 与 β_j 之间的交互作用的效应, ε_{ijk} 是随机误差分量. 两个因子最初都假定是**固定**的, 处理效应规定为与总平均的偏差, 所以 $\sum\limits_{i=1}^{a}\tau_i=0$, $\sum\limits_{j=1}^{b}\beta_j=0$. 同样, 交互作用效应是固定的并限定它满足 $\sum\limits_{i=1}^{a}(\tau\beta)_{ij}=\sum\limits_{j=1}^{b}(\tau\beta)_{ij}=0$. 因为实验有 n 次重复, 所以共有 abn 个观测值.

析因设计的另一个可能的模型是**均值模型**:

$$y_{ijk}=\mu_{ij}+\varepsilon_{ijk}\quad\begin{cases}i=1,2,\cdots,a\\ j=1,2,\cdots,b\\ k=1,2,\cdots,n\end{cases}$$

其中第 ij 个单元的均值为

$$\mu_{ij}=\mu+\tau_i+\beta_j+(\tau\beta)_{ij}$$

我们也可以用 5.1 节中的**回归模型**. 回归模型对实验中有一个或多个定量因子时特别有用. 本章主要利用效应模型 [(5.1) 式], 5.5 节中用回归模型.

二因子析因设计中, 行因子 A 与列因子 B(或处理) 同样重要. 具体来说, 我们想要检验行处理效应的等式假设, 即

$$H_0:\ \tau_1=\tau_2=\cdots=\tau_a=0,\qquad H_1:\ 至少一个\ \tau_i\neq 0\tag{5.2a}$$

以及列处理效应的等式假设, 即

$$H_0:\ \beta_1=\beta_2=\cdots=\beta_b=0,\qquad H_1:\ 至少一个\ \beta_j\neq 0 \tag{5.2b}$$

我们还想判定行与列处理究竟有没有交互作用. 于是, 我们亦想检验

$$H_0:\ (\tau\beta)_{ij}=0,\ 对所有的\ i,j,\qquad H_1:\ 至少一个(\tau\beta)_{ij}\neq 0 \tag{5.2c}$$

现在讨论如何用**二因子方差分析**来检验这些假设.

5.3.2 固定效应模型的统计分析

令 $y_{i..}$ 表示在因子 A 在第 i 水平下所有观测值的总和, $y_{.j.}$ 表示在因子 B 在第 j 水平下所有观测值的总和, $y_{ij.}$ 表示第 ij 单元中所有观测值的总和, $y_{...}$ 表示全部观测值的总和, 对应于行、列、单元与总和的平均值分别记为 $\bar{y}_{i..}$, $\bar{y}_{.j.}$, $\bar{y}_{ij.}$ 与 $\bar{y}_{...}$. 数学表达式为:

$$\begin{aligned}
&y_{i..}=\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk} && \bar{y}_{i..}=\frac{y_{i..}}{bn} && i=1,2,\cdots,a\\
&y_{.j.}=\sum_{i=1}^{a}\sum_{k=1}^{n}y_{ijk} && \bar{y}_{.j.}=\frac{y_{.j.}}{an} && j=1,2,\cdots,b\\
&y_{ij.}=\sum_{k=1}^{n}y_{ijk} && \bar{y}_{ij.}=\frac{y_{ij.}}{n} && \begin{matrix} i=1,2,,\cdots,a\\ j=1,2,\cdots,b\end{matrix}\\
&y_{...}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk} && \bar{y}_{...}=\frac{y_{...}}{abn} &&
\end{aligned} \tag{5.3}$$

总校正平方和可写为

$$\begin{aligned}
\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{...})^2 &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}[(\bar{y}_{i..}-\bar{y}_{...})+(\bar{y}_{.j.}-\bar{y}_{...})\\
&\quad+(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})+(y_{ijk}-\bar{y}_{ij.})]^2\\
&= bn\sum_{i=1}^{a}(\bar{y}_{i..}-\bar{y}_{...})^2+an\sum_{j=1}^{b}(\bar{y}_{.j.}-\bar{y}_{...})^2\\
&\quad+n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2\\
&\quad+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{ij.})^2
\end{aligned} \tag{5.4}$$

这是因为右边 6 项交叉乘积为零. 我们注意到, 总平方和已分解为 "行" 即因子 A 的平方和 (SS_A), "列" 即因子 B 的平方和 (SS_B), A 与 B 间的交互作用的平方和 (SS_{AB}), 以及起因于误差的平方和 (SS_E). 这是二因子析因设计的基本 ANOVA 等式, 由 (5.4) 式右端项最后的分量可以看出, 要得出误差平方和至少要有两次重复 ($n\geqslant 2$).

用符号记 (5.4) 式为

$$SS_T=SS_A+SS_B+SS_{AB}+SS_E \tag{5.5}$$

各个平方和的自由度是

效应	自由度
A	$a-1$
B	$b-1$
AB 交互作用	$(a-1)(b-1)$
误差	$ab(n-1)$
总和	$abn-1$

对平方和的 $abn-1$ 个总的自由度的分配, 验证如下: 主效应 A 与主效应 B 分别有 a 个水平与 b 个水平, 因此, 它们有 $a-1$ 个自由度与 $b-1$ 个自由度; 交互作用的自由度是单元的自由度 (为 $ab-1$) 减去两个主效应 A 与 B 的自由度, 即 $ab-1-(a-1)-(b-1)=(a-1)(b-1)$; 在 ab 个单元的每一单元内, n 次重复间有 $n-1$ 个自由度, 于是, 有 $ab(n-1)$ 个误差自由度. (5.5) 式右端项的自由度加起来就是总的自由度.

每一平方和除以它的自由度就是**均方**. 均方的期望值是

$$\mathrm{E}(MS_A)=\mathrm{E}\left(\frac{SS_A}{a-1}\right)=\sigma^2+\frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$$

$$\mathrm{E}(MS_B)=\mathrm{E}\left(\frac{SS_B}{b-1}\right)=\sigma^2+\frac{an\sum_{j=1}^{b}\beta_j^2}{b-1}$$

$$\mathrm{E}(MS_{AB})=\mathrm{E}\left(\frac{SS_{AB}}{(a-1)(b-1)}\right)=\sigma^2+\frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$$

$$\mathrm{E}(MS_E)=\mathrm{E}\left(\frac{SS_E}{ab(n-1)}\right)=\sigma^2$$

如果没有行处理效应、没有列处理效应以及没有交互作用的零假设为真, 则 MS_A、MS_B、MS_{AB} 与 MS_E 都估计了 σ^2. 然而, 比方说, 只要行处理效应之间有差异, 则 MS_A 将大于 MS_E. 同理, 当有列处理效应或出现交互作用时, 则对应的均方将大于 MS_E. 因此, 为了检验主效应与交互作用的显著性, 将对应的均方除以误差均方. 这个比值较大表示所得的数据不支持零假设.

如果模型 [(5.1) 式] 是合适的, 误差项 ε_{ijk} 服从正态独立分布且有常值方差 σ^2, 则均方的每一比值 MS_A/MS_E、MS_B/MS_E 以及 MS_{AB}/MS_E 都服从 F 分布, 其分子的自由度分别是 $a-1$、$b-1$ 与 $(a-1)(b-1)$, 分母的自由度是 $ab(n-1)$①, 而临界区域就是 F 分布的上尾部, 检验方法通常概括在一张**方差分析表**中, 如表 5.3 所示.

我们一般都用统计软件包进行方差分析计算. 但是, 也可直接手工计算 (5.5) 式中的平方和. 可以将方差分析恒等式中的每一元素写成

$$y_{ijk}-\bar{y}_{...}=(\bar{y}_{i..}-\bar{y}_{...})+(\bar{y}_{.j.}-\bar{y}_{...})+(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})+(y_{ijk}-\bar{y}_{ij.})$$

并在工作表的列中计算它们. 然后对每列平方并求和得到 ANOVA 的平方和. 也可以利用基于行、列以及单元和的计算公式. 总的平方和通常由

$$SS_T=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}^2-\frac{y_{...}^2}{abn} \tag{5.6}$$

① 如前所述, F 检验法亦可看作为随机化检验的近似方法.

表 5.3　二因子析因设计的方差分析表，固定效应模型

方差来源	平方和	自由度	均　方	F_0
处理 A	SS_A	$a-1$	$MS_A=\frac{SS_A}{a-1}$	$F_0=\frac{MS_A}{MS_E}$
处理 B	SS_B	$b-1$	$MS_B=\frac{SS_B}{b-1}$	$F_0=\frac{MS_B}{MS_E}$
交互作用	SS_{AB}	$(a-1)(b-1)$	$MS_{AB}=\frac{SS_{AB}}{(a-1)(b-1)}$	$F_0=\frac{MS_{AB}}{MS_E}$
误差	SS_E	$ab(n-1)$	$MS_E=\frac{SS_E}{ab(n-1)}$	
总和	SS_T	$abn-1$		

计算出来. 主效应平方和是

$$SS_A=\frac{1}{bn}\sum_{i=1}^{a}y_{i..}^2-\frac{y_{...}^2}{abn} \tag{5.7}$$

和

$$SS_B=\frac{1}{an}\sum_{j=1}^{b}y_{.j.}^2-\frac{y_{...}^2}{abn} \tag{5.8}$$

分两步得出 SS_{AB} 是方便的. 首先计算 ab 个单元总和间的平方和, 称为 “小计” 平方和：

$$SS_{\text{小计}}=\frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij.}^2-\frac{y_{...}^2}{abn}$$

这一平方和还包含 SS_A 与 SS_B. 因此, 第二步是计算 SS_{AB}：

$$SS_{AB}=SS_{\text{小计}}-SS_A-SS_B \tag{5.9}$$

用减法计算 SS_E 得

$$SS_E=SS_T-SS_{AB}-SS_A-SS_B \tag{5.10}$$

即

$$SS_E=SS_T-SS_{\text{小计}}$$

例5.1　电池设计实验

表 5.4 表示 5.3.1 节所述的电池设计例子中所观测到的有效寿命 (以小时计). 行与列的总和在表的边上, 单元总和用圆圈圈起来.

表 5.4　电池设计例子的寿命数据 (单位：小时)

材料类型	温度 (°F)									
	15			70			125			$y_{i..}$
1	130, 74	155, 180	(539)	34, 80	40, 75	(229)	20, 82	70, 58	(230)	998
2	150, 159	188, 126	(623)	136, 106	122, 115	(479)	25, 58	70, 45	(198)	1 300
3	138, 168	110, 160	(576)	174, 150	120, 139	(583)	96, 82	104, 60	(342)	1 501
$y_{.j.}$	1 738			1 291			770			3 799=$y_{...}$

用 (5.6) 式至 (5.10) 式算得平方和如下：

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

$$= (130)^2 + (155)^2 + (74)^2 + \cdots + (60)^2 - \frac{(3\ 799)^2}{36} = 77\ 646.97$$

$$SS_{材料} = \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - \frac{y_{...}^2}{abn}$$

$$= \frac{1}{(3)(4)}[(998)^2 + (1\ 300)^2 + (1\ 501)^2] - \frac{(3\ 799)^2}{36} = 10\ 683.72$$

$$SS_{温度} = \frac{1}{an}\sum_{j=1}^{b} y_{.j.}^2 - \frac{y_{...}^2}{abn}$$

$$= \frac{1}{(3)(4)}[(1\ 738)^2 + (1\ 291)^2 + (770)^2] - \frac{(3\ 799)^2}{36} = 39\ 118.72$$

$$SS_{交互作用} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{...}^2}{abn} - SS_{材料} - SS_{温度}$$

$$= \frac{1}{4}[(539)^2 + (229)^2 + \cdots + (342)^2] - \frac{(3\ 799)^2}{36} - 10\ 683.72 - 39\ 118.72 = 9\ 613.78$$

$$SS_E = SS_T - SS_{材料} - SS_{温度} - SS_{交互作用}$$

$$= 77\ 646.97 - 10\ 683.72 - 39\ 118.72 - 9\ 613.78 = 18\ 230.75$$

方差分析如表 5.5 所示. 因为 $F_{0.05,4,27} = 2.73$, 所以结论是: 材料类型与温度之间有显著的交互作用. 又因 $F_{0.05,2,27} = 3.35$, 所以材料类型与温度的主效应也是显著的. 表 5.5 也给出了检验统计量的 P 值.

表 5.5 电池寿命数据的方差分析表

方差来源	平方和	自由度	均 方	F_0	P 值
材料类型	10 683.72	2	5 341.86	7.91	0.002 0
温度	39 118.72	2	19 559.36	28.97	0.000 1
交互作用	9 613.78	4	2 403.44	3.56	0.018 6
误差	18 230.75	27	675.21		
总和	77 646.97	35			

为解释这一实验结果, 可构造每一处理组合的平均响应图, 见图 5.9. 显著性交互作用由线段的不平行性表明. 一般说来, 不管是什么材料, 在低温处的寿命都较长. 从低温变化至中等温度时, 用材料类型 3 生

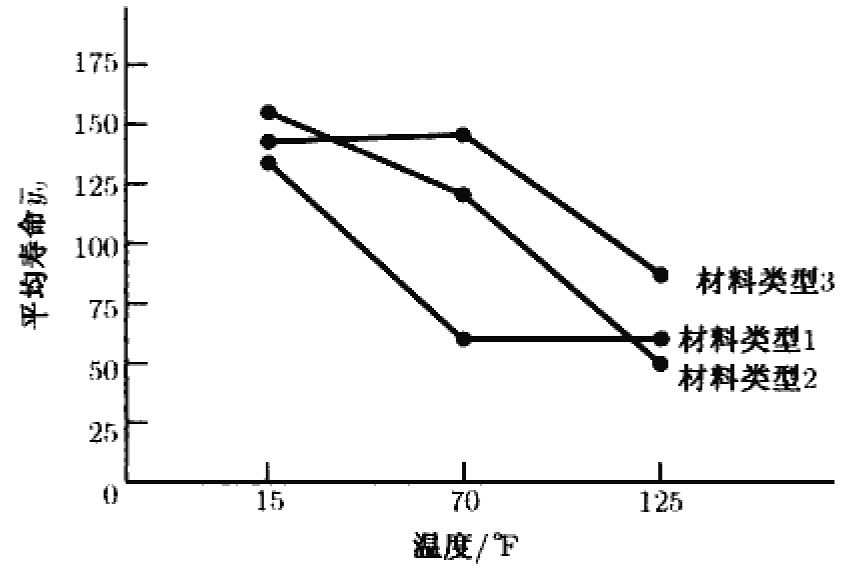


图 5.9 例 5.1 材料类型—温度的图

产的电池的寿命实际上是增加的, 而对类型 1 与类型 2 来说则是减少的. 从中等温度变至高温时, 用材料类型 2 与类型 3 生产的电池的寿命减少, 而由类型 1 生产的电池的寿命不变. 因此材料类型 3 给出最好的结果, 如果我们希望尽量减少温度变化时电池寿命的损失.

1. 多重比较

当方差分析表明行均值或列均值有差异时, 通常人们乐意做各个行均值与列均值之间的比较, 以便发现具体的差别. 第 3 章讨论的多重比较法在此处有用.

今说明 Tukey 检验法在例 5.1 的电池寿命数据中的用途. 在这一实验中, 交互作用是显著的. 当交互作用显著时, 一个因子 (例如, A) 的均值间的比较可能由于 AB 的交互作用而模糊不清. 解决这类问题的一种方法是, 将因子 B 固定在一特定水平上, 在此水平上对因子 A 的均值应用 Tukey 检验法. 为说明起见, 设在例 5.1 中我们感兴趣于检验 3 种材料类型的均值差. 因交互作用显著, 我们仅在一个温度水平 [比方说水平 2(70 ℉)] 上做这种比较. 设误差方差的最好估计是方差分析表中的 MS_E, 此处利用了实验的误差方差对所有处理组合都是相同的这一假定.

这 3 种材料类型在 70 ℉ 时的均值依递增顺序排列是

$$\bar{y}_{12.} = 57.25 \quad (\text{材料类型 } 1)$$
$$\bar{y}_{22.} = 119.75 \quad (\text{材料类型 } 2)$$
$$\bar{y}_{32.} = 145.75 \quad (\text{材料类型 } 3)$$

因此

$$T_{0.05} = q_{0.05}(3,27)\sqrt{\frac{MS_E}{n}} = 3.50\sqrt{\frac{675.21}{4}} = 45.47$$

其中 $q_{0.05}(3,27) \simeq 3.50$ 可以通过附表Ⅷ的插值得到. 配对比较得

$$3\text{ 对 }1 \quad 145.75 - 57.25 = 88.50 > T_{0.05} = 45.47$$
$$3\text{ 对 }2 \quad 145.75 - 119.75 = 26.00 < T_{0.05} = 45.47$$
$$2\text{ 对 }1 \quad 119.75 - 57.25 = 62.50 > T_{0.05} = 45.47$$

这一分析表明, 当温度水平为 70 ℉ 时, 材料类型 2 与类型 3 的电池平均寿命是相同的, 而材料类型 1 的电池平均寿命显著地低于材料类型 2 与类型 3.

当交互作用显著时, 实验者可以比较所有 ab 个单元的均值, 以便确定哪一些有显著性差异. 在这一分析中, 单元均值间的差异既包含了交互作用效应的又包含了所有主效应的. 在例 5.1 中, 9 个单元均值的所有可能的配对之间会得出 36 个比较.

2. 计算机输出

对于例 5.1 的电池寿命数据, 图 5.10 给出了 Design-Expert 的主要计算输出. 注意到

$$SS_{\text{模型}} = SS_{\text{材料}} + SS_{\text{温度}} + SS_{\text{交互作用}} = 10\ 683.72 + 39\ 118.72 + 9\ 613.78 = 59\ 416.22$$

它有 8 个自由度. F 检验被列出, 以便对模型的变异来源进行分析. 因为 P 值较小 (<0.000 1), 所以, 该检验的解释是, 模型的 3 项中至少有一项是显著的. 后面对单个模型项 (A, B, AB) 进行检验. 又

$$R^2 = \frac{SS_{\text{模型}}}{SS_{\text{总和}}} = \frac{59\ 416.22}{77\ 646.97} = 0.765\ 2$$

Response: Life in hours

ANOVA for Selected Factorial Model

Analysis of variance table[Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Model	59416.22	8	7427.03	11.00	<0.0001	significant
A	10683.72	2	5341.86	7.91	0.0020	
B	39118.72	2	19559.36	28.97	<0.0001	
AB	9613.78	4	2403.44	3.56	0.0186	
Residual	18230.75	27	675.21			
Lack of Fit	0.000	0				
Pure Error	18230.75	27	675.21			
Cor Total	77646.97	35				

Std.Dev.	25.98	R-Squared	0.7652
Mean	105.53	Adj R-Squared	0.6956
C.V.	24.62	Pred R-Squared	0.5826
PRESS	32410.22	Adeq Precision	8.178

Diagnostics Case Statistics

Standard Order	Actual Value	Predicted Value	Residual	Leverage	Student Residual	Cook's Distance	Outlier *t*
1	130.00	134.75	−4.75	0.250	−0.211	0.002	−0.207
2	74.00	134.75	−60.75	0.250	−2.700	0.270	−3.100
3	155.00	134.75	20.25	0.250	0.900	0.030	0.897
4	180.00	134.75	45.25	0.250	2.011	0.150	2.140
5	150.00	155.75	−5.75	0.250	−0.256	0.002	−0.251
6	159.00	155.75	3.25	0.250	0.144	0.001	0.142
7	188.00	155.75	32.25	0.250	1.433	0.076	1.463
8	126.00	155.75	−29.75	0.250	−1.322	0.065	−1.341
9	138.00	144.00	26.00	0.250	−0.267	0.003	−0.262
10	168.00	144.00	24.00	0.250	1.066	0.042	1.069
11	110.00	144.00	−34.00	0.250	−1.511	0.085	−1.550
12	160.00	144.00	16.00	0.250	0.711	0.019	0.704
13	34.00	57.25	−23.25	0.250	−1.033	0.040	−1.035
14	80.00	57.25	22.75	0.250	1.011	0.038	1.011
15	40.00	57.25	−17.25	0.250	−0.767	0.022	−0.761
16	75.00	57.25	17.75	0.250	0.789	0.023	0.783
17	136.00	119.75	16.25	0.250	0.722	0.019	0.716
18	106.00	119.75	−13.75	0.250	−0.611	0.014	−0.604
19	122.00	119.75	2.25	0.250	0.100	0.000	0.098
20	115.00	119.75	−4.75	0.250	−0.211	0.002	−0.207
21	174.00	145.75	28.25	0.250	1.255	0.058	1.269
22	150.00	145.75	4.25	0.250	0.189	0.001	0.185
23	120.00	145.75	−25.75	0.250	−1.144	0.048	−1.151
24	139.00	145.75	−6.75	0.250	−0.300	0.003	−0.295
25	20.00	57.50	−37.50	0.250	−1.666	0.103	−1.726
26	82.00	57.50	24.50	0.250	1.089	0.044	1.093
27	70.00	57.50	12.50	0.250	0.555	0.011	0.548
28	58.00	57.50	0.50	0.250	0.022	0.000	0.022
29	25.00	49.50	−24.50	0.250	−1.089	0.044	−1.093
30	58.00	49.50	8.50	0.250	0.378	0.005	0.372
31	70.00	49.50	20.50	0.250	0.911	0.031	0.908
32	45.00	49.50	−4.50	0.250	−0.200	0.001	−0.196
33	96.00	85.50	10.50	0.250	0.467	0.008	0.460
34	82.00	85.50	−3.50	0.250	−0.156	0.001	−0.153
35	104.00	85.50	18.50	0.250	0.822	0.025	0.817
36	60.00	85.50	−25.50	0.250	−1.133	0.048	−1.139

图 5.10 例 5.1 的 Design-Expert 输出

也就是说, 电池寿命中的变异性的大约 77%可以用电池的板极材料、温度, 以及材料类型 — 温度的交互作用来解释. 拟合模型的残差也显示在计算输出中. 现在, 我们说明在模型适合性检验中怎样利用这些残差.

5.3.3　模型适合性检验

在采用方差分析的结论之前, 应该检验设定模型的适合性. 和以前一样, 基本的诊断工具是**残差分析**. 二因子析因模型的残差是

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk} \tag{5.11}$$

又因拟合值 $\hat{y}_{ijk} = \bar{y}_{ij.}$(第 ij 个单元中观测值的平均值), (5.11) 式变为

$$e_{ijk} = y_{ijk} - \bar{y}_{ij.} \tag{5.12}$$

例 5.1 中电池寿命数据的残差显示在 Design–Expert 计算输出和表 5.6 里. 这些残差的正态概率图 (图 5.11) 并未显出任何特别的麻烦, 尽管最大的负残差 (−60.75, 材料类型 1 在 15℉ 处) 确实与其他残差稍有不同. 这一残差的标准化值是 $-60.75/\sqrt{675.21} = -2.34$, 这仅是绝对值大于 2 的一个残差.

表 5.6　例 5.1 的残差

材料类型	温度 (℉)					
	15		70		125	
1	−4.75	20.25	−23.25	−17.25	−37.50	12.50
	−60.75	45.25	22.75	17.75	24.50	0.50
2	−5.75	32.25	16.25	2.25	−24.50	20.50
	3.25	−29.75	−13.75	−4.75	8.50	−4.50
3	−6.00	−34.00	28.25	−25.75	10.50	18.50
	24.00	16.00	4.25	−6.75	−3.50	−25.50

图 5.12 作出了残差与拟合值 $\hat{y}_{ijk}$ 的关系图. 此图表明, 当电池寿命增加时, 残差方差有微弱的增长趋势. 图 5.13 与图 5.14 分别是残差与材料类型的关系图和残差与温度的关系图.

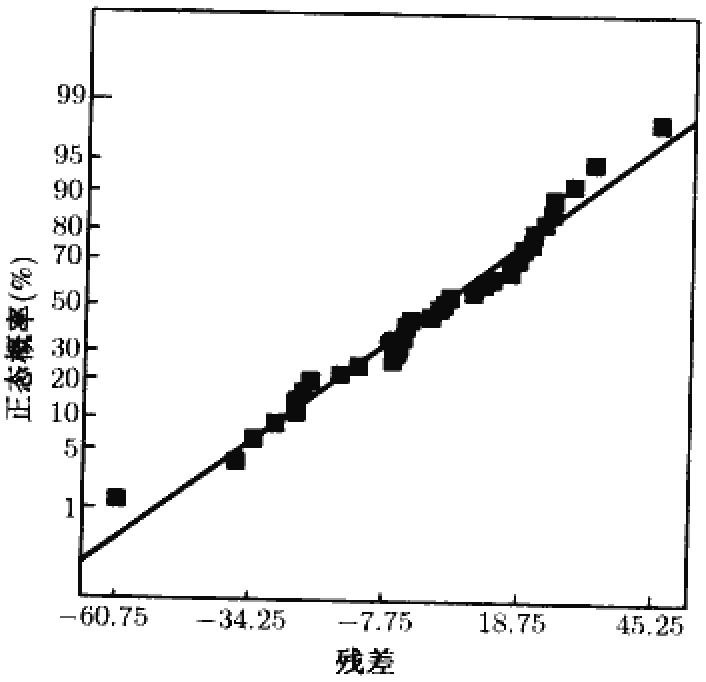


图 5.11　例 5.1 的残差与正态概率的关系图

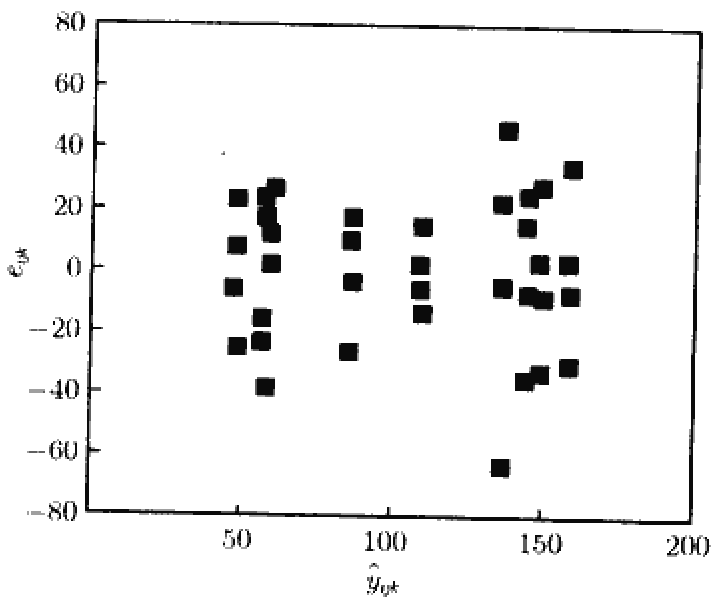


图 5.12　例 5.1 的残差与拟合值 $\hat{y}_{ijk}$ 的关系图

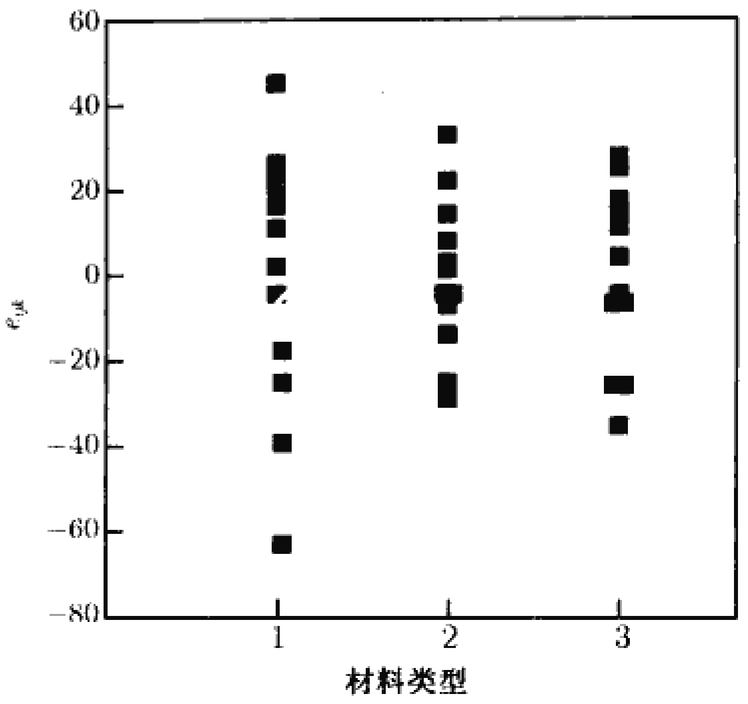


图 5.13 例 5.1 的残差与材料类型的关系图

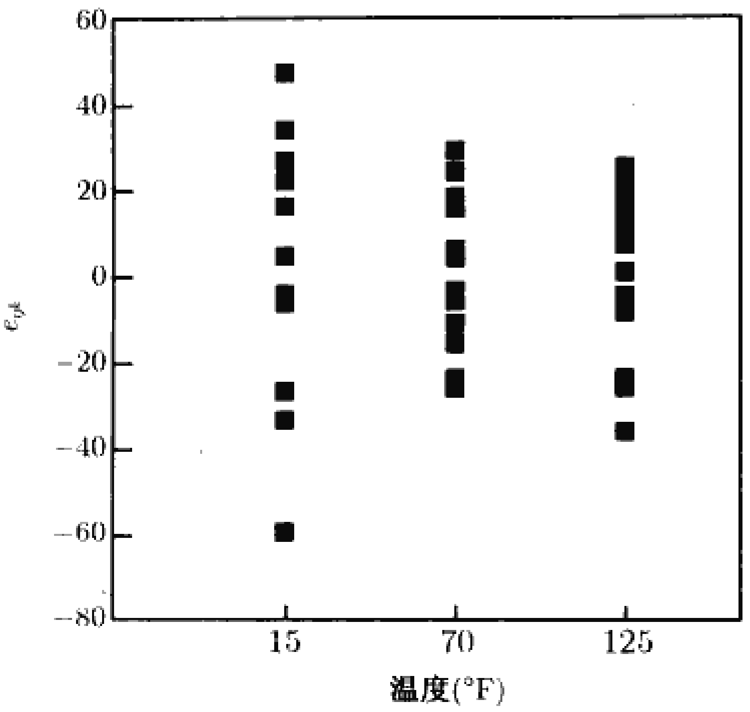


图 5.14 例 5.1 的残差与温度的关系图

这两张图都显示出方差的微弱不等性, 15 °F 与材料类型 1 的处理组合, 比之其他情况可能有较大的方差.

从表 5.6 中看出, 在 15 °F — 材料类型 1 这一单元中, 包含了两个极端残差值 (−60.75 与 45.25). 这两个残差值对于由图 5.12、图 5.13 与图 5.14 检验出的方差不等式起了主要作用. 重新审查这些数据, 并未显出任何明显的问题, 比如记录错误等, 所以我们把这些响应作为真实数据接受下来. 这一特别的处理组合比其他处理组合有可能产生稍为不稳定的电池寿命. 然而, 问题不至于严重到对分析和结论产生严重影响的程度.

5.3.4 估计模型参数

二因子析因设计的效应模型

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \tag{5.13}$$

的参数可以用最小二乘法来估计. 因为这一模型有 $1+a+b+ab$ 个参数需要估计, 所以有 $1+a+b+ab$ 个正规方程. 用 3.9 节的方法, 不难证明, 正规方程组是

$$\mu : abn\hat{\mu} + bn\sum_{i=1}^{a}\hat{\tau}_i + an\sum_{j=1}^{b}\hat{\beta}_j + n\sum_{i=1}^{a}\sum_{j=1}^{b}(\widehat{\tau\beta})_{ij} = y_{\cdots} \tag{5.14a}$$

$$\tau_i : bn\hat{\mu} + bn\hat{\tau}_i + n\sum_{j=1}^{b}\hat{\beta}_j + n\sum_{j=1}^{b}(\widehat{\tau\beta})_{ij} = y_{i\cdot\cdot} \quad i = 1, 2, \cdots, a \tag{5.14b}$$

$$\beta_j : an\hat{\mu} + n\sum_{i=1}^{a}\hat{\tau}_i + an\hat{\beta}_j + n\sum_{i=1}^{a}(\widehat{\tau\beta})_{ij} = y_{\cdot j\cdot} \quad j = 1, 2, \cdots, b \tag{5.14c}$$

$$(\tau\beta)_{ij} : n\hat{\mu} + n\hat{\tau}_i + n\hat{\beta}_j + n(\widehat{\tau\beta})_{ij} = y_{\cdot j\cdot} \quad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b \end{cases} \tag{5.14d}$$

为方便起见, 对应于每一正规方程的参数都已在 (5.14) 式的左边表示出来.

效应模型 [(5.13) 式] 是一个超参数模型. 注意到将 (5.14b) 式中的 a 个方程加起来就是 (5.14a) 式, (5.14c) 式中的 b 个方程加起来就是 (5.14a) 式. 同理, 对任一 i, 将 (5.14d) 式关于

j 加起来就是 (5.14b) 式; 对任一 j, 将 (5.14d) 式关于 i 加起来就是 (5.14c) 式. 因此, 在此方程组中, 有 $a+b+1$ 个**线性相关**, 从而不存在唯一解. 为求得一个解, 我们加进约束

$$\sum_{i=1}^{a}\hat{\tau}_i=0 \tag{5.15a}$$

$$\sum_{j=1}^{b}\hat{\beta}_j=0 \tag{5.15b}$$

$$\sum_{i=1}^{a}(\widehat{\tau\beta})_{ij}=0 \quad j=1,2,\cdots,b \tag{5.15c}$$

$$\sum_{j=1}^{b}(\widehat{\tau\beta})_{ij}=0 \quad j=1,2,\cdots,a \tag{5.15d}$$

(5.15a) 式与 (5.15b) 式构成两个约束, 而 (5.15c) 式与 (5.15d) 式形成 $a+b-1$ 个独立的约束. 因此, 总共有 $a+b+1$ 个约束, 恰是所需的数目.

应用这些约束, 还大大简化了正规方程组 [(5.14) 式], 求得解为

$$\begin{aligned}
\hat{\mu}&=\bar{y}_{...}\\
\hat{\tau}_i&=\bar{y}_{i..}-\bar{y}_{...} \quad i=1,2,\cdots,a\\
\hat{\beta}_j&=\bar{y}_{.j.}-\bar{y}_{...} \quad j=1,2,\cdots,b\\
(\widehat{\tau\beta})_{ij}&=\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...} \quad \begin{cases} i=1,2,\cdots,a\\ j=1,2,\cdots,b\end{cases}
\end{aligned} \tag{5.16}$$

正规方程组的这一解有很直观的意义. 行处理效应估计为行平均减去总平均, 列处理效应估计为列平均减去总平均, 第 ij 个交互作用估计为第 ij 个单元的平均减去总平均、第 i 个行效应和第 j 个列效应.

用 (5.16) 式, 可求得 y_{ijk} 的**拟合值**为

$$\hat{y}_{ijk}=\hat{\mu}+\hat{\tau}_i+\hat{\beta}_j+(\widehat{\tau\beta})_{ij}=\bar{y}_{...}+(\bar{y}_{i..}-\bar{y}_{...})+(\bar{y}_{.j.}-\bar{y}_{...})+(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})=\bar{y}_{ij.}$$

也就是说, 第 ij 个单元中的第 k 个观测值估计为在那一单元中 n 个观测值的平均. 这一结果曾在 (5.12) 式中用来求二因子析因模型的残差.

因为解正规方程组时用了约束 [(5.15) 式], 所以模型参数的估计量不是唯一的. 不过, 模型参数的某些重要函数是可估的, 即不管约束怎样选取, 它们可以唯一地被估计. 例如 $\tau_i-\tau_u+\overline{(\tau\beta)}_{i.}-\overline{(\tau\beta)}_{u.}$, 它可看作为因子 A 的第 i 个与第 u 个水平的 “真实” 差. 注意, 任一主效应的水平间的真实差都包含 “平均” 交互作用效应. 如同前面已说明的, 正是这个结果在出现交互作用时它会干扰对主效应的检验. 一般说来, 当模型参数的任一函数是正规方程组左端项的线性组合时, 必是可估的. 这一性质在第 3 章讨论单因子模型时也给予了说明. 更多信息见本章的补充材料.

5.3.5 样本量的选择

可以利用附录图 V 的抽检特性曲线, 帮助实验者给二因子析因设计确定一个恰当的样本量 (重复次数, n). 参数 Φ^2 的恰当值以及分子和分母的自由度如表 5.7 所示.

表 5.7 关于固定效应模型二因子析因设计，附录图 V 的抽检特性曲线参数

因子	Φ^2	分子自由度	分母自由度
A	$\dfrac{bn\sum_{i=1}^{a}\tau_i^2}{a\sigma^2}$	$a-1$	$ab(n-1)$
B	$\dfrac{an\sum_{j=1}^{b}\beta_j^2}{b\sigma^2}$	$b-1$	$ab(n-1)$
AB	$\dfrac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{\sigma^2[(a-1)(b-1)+1]}$	$(a-1)(b-1)$	$ab(n-1)$

利用这些曲线的一种很有效的方法是, 寻找与任意两个处理的特定的均值差相对应的 Φ^2 的最小值. 例如, 当任意两行的均值差为 D 时, 则 Φ^2 的最小值是

$$\Phi^2=\frac{nbD^2}{2a\sigma^2} \tag{5.17}$$

而当任意两列的均值差为 D 时, 则 Φ^2 的最小值是

$$\Phi^2=\frac{naD^2}{2b\sigma^2} \tag{5.18}$$

当任意两个交互作用效应之差为 D 时, 相应的 Φ^2 的最小值是

$$\Phi^2=\frac{nD^2}{2\sigma^2[(a-1)(b-1)+1]} \tag{5.19}$$

用例 5.1 的电池寿命数据来说明这些等式的用途. 设在实验之前, 我们确定, 如果任意两种温度对应的电池寿命的均值差大到 40 小时的话, 则以高概率拒绝零假设. 于是差 $D=40$ 有工程显著性, 如果再假定电池寿命的标准差近似为 25, 则由 (5.18) 式得

$$\Phi^2=\frac{naD^2}{2b\sigma^2}=\frac{n(3)(40)^2}{2(3)(25)^2}=1.28n$$

它给出了 Φ^2 的最小值. 取 $\alpha=0.05$, 用附录图 V 得下表:

n	Φ^2	Φ	ν_1= 分子自由度	ν_2= 误差自由度	β
2	2.56	1.60	2	9	0.45
3	3.84	1.96	2	18	0.18
4	5.12	2.26	2	27	0.06

注意到 n= 4 次重复时, 得出的风险 β 约为 0.06. 或者说, 对应于任意两种温度水平, 电池寿命的均值差大到 40 小时的话, 则以大约 94%的机会拒绝零假设. 这样一来, 结论是, 只要我们对电池寿命的标准差的估计出入不太大, 那么做 4 次重复足以达到所要求的灵敏度. 如果拿不准的话, 实验者可以用其他的 σ 值来重复上述程序, 以便确定错估该参数对设计的灵敏度的影响.

5.3.6 假定在二因子模型中没有交互作用

有时候, 实验者觉得, **没有交互作用的二因子模型**是恰当的, 即

$$y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk} \quad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b \\ k = 1, 2, \cdots, n \end{cases} \tag{5.20}$$

但是我们必须对在模型中舍弃交互作用项抱十分谨慎的态度, 因为显著的交互作用的实际存在会对数据的解释产生极大的冲击.

无交互作用的二因子析因模型的统计分析是简单的. 用无交互作用模型 [(5.20) 式] 对例 5.1 的电池寿命数据所作的分析见表 5.8. 与先前的分析一样, 两个主效应是显著的. 然而, 一旦对这些数据进行残差分析, 就立即看出, 无交互作用模型是不合适的. 对无交互作用的二因子模型来说, 拟合值是 $\hat{y}_{ijk} = \bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...}$. $\bar{y}_{ij.} - \hat{y}_{ij}$(单元平均值减该单元的拟合值) 与拟合值 $\hat{y}_{ijk}$ 的关系图如图 5.15 所示. 现在, 量 $\bar{y}_{ij.} - \hat{y}_{ijk}$ 可看作为: 在无交互作用的假定下, 被观测的单元均值与该单元的均值估计之差. 这些量的任何特定的模式都暗示着交互作用的存在. 图 5.15 显示了一种明显的模式, 因为量 $\bar{y}_{ij.} - \hat{y}_{ijk}$ 从正到负再到正再到负. 这种结构是材料类型与温度之间交互作用的结果.

表 5.8　假定无交互作用时电池寿命数据的方差分析

方差来源	平方和	自由度	均方	F_0
材料类型	10 683.72	2	5 341.86	5.95
温度	39 118.72	2	19 559.36	21.78
误差	27 844.52	31	898.21	
总和	77 646.96	35		

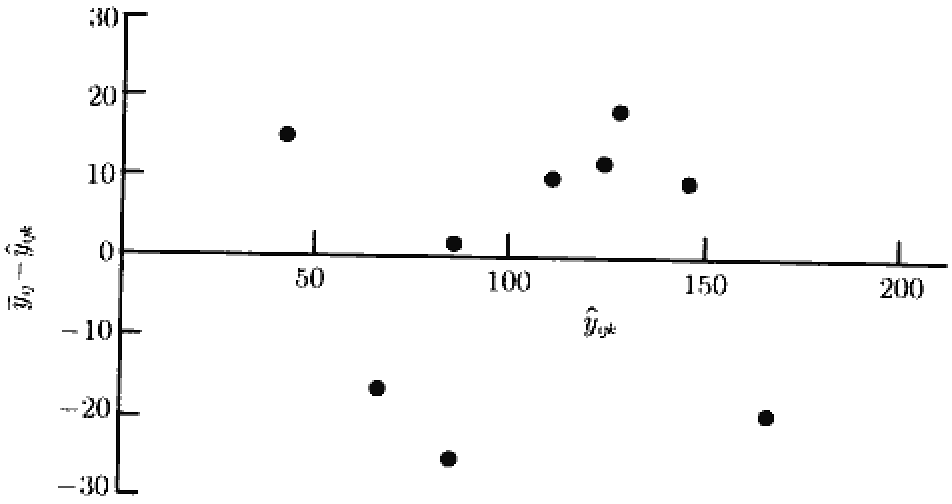


图 5.15　$\bar{y}_{ij.} - \hat{y}_{ijk}$ 与 $\hat{y}_{ijk}$(电池寿命数据) 的关系图

5.3.7　每单元一个观测值

有时, 人们会遇到只做一次重复的二因子实验, 也就是说, 每单元只有一个观测值. 如果有两个因子且每单元只有一个观测值, 则效应模型是

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b \end{cases} \tag{5.21}$$

假设两个因子固定, 此时的方差分析如表 5.9 所示.

由期望均方可见, 不能估计误差方差 σ^2. 也就是说, 二因子交互效应 $(\tau\beta)_{ij}$ 与实验误差不能以任何明显的方式分离开来. 因此, 除非交互作用效应为零, 否则, 就没有关于主效应的检验法. 如果没有交互作用存在, 则对一切 i 和 j, $(\tau\beta)_{ij}=0$, 似乎可取的模型是

$$y_{ij}=\mu+\tau_i+\beta_j+\varepsilon_{ij} \quad \begin{cases} i=1,2,\cdots,a \\ j=1,2,\cdots,b \end{cases} \tag{5.22}$$

如果这一模型 [(5.22) 式] 是恰当的, 则表 5.9 的残差均方就是 σ^2 的一个无偏估计量, 而且将 MS_A、MS_B 与 $MS_{残差}$ 进行比较, 就可以检验主效应.

表 5.9 每单元一个观测值时, 二因子模型的方差分析

方差来源	平方和	自由度	均　方	期望均方
行 (A)	$\sum_{i=1}^{a}\frac{y_{i.}^2}{b}-\frac{y_{..}^2}{ab}$	$a-1$	MS_A	$\sigma^2+\frac{b\sum\tau_i^2}{a-1}$
列 (B)	$\sum_{j=1}^{b}\frac{y_{.j}^2}{a}-\frac{y_{..}^2}{ab}$	$b-1$	MS_B	$\sigma^2+\frac{a\sum\beta_j^2}{b-1}$
残差即 AB	减法	$(a-1)(b-1)$	$MS_{残差}$	$\sigma^2+\frac{\sum\sum(\tau\beta)_{ij}^2}{(a-1)(b-1)}$
总和	$\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij}^2-\frac{y_{..}^2}{ab}$	$ab-1$		

在确定交互作用是否存在方面, 可以使用 Tukey(1949a) 提出的检验法. 此法假定交互作用项以特别简单的形式出现, 即

$$(\tau\beta)_{ij}=\gamma\tau_i\beta_j$$

其中 γ 是未知常数. 利用以这种方式定义交互作用项, 就可以使用回归方法来检验交互作用项的显著性. 这一检验法将方差的残差平方和分解为属于非可加性 (交互作用) 的单自由度分量和自由度为 $(a-1)(b-1)-1$ 的误差分量. 由计算得:

$$SS_N=\frac{\left[\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij}y_{i.}y_{.j}-y_{..}\left(SS_A+SS_B+\frac{y_{..}^2}{ab}\right)\right]^2}{abSS_ASS_B} \tag{5.23}$$

它有 1 个自由度; 而

$$SS_{误差}=SS_{残差}-SS_N \tag{5.24}$$

它有 $(a-1)(b-1)-1$ 个自由度. 为了检验交互作用是否存在, 计算

$$F_0=\frac{SS_N}{SS_{误差}/[(a-1)(b-1)-1]} \tag{5.25}$$

当 $F_0>F_{\alpha,1,(a-1)(b-1)-1}$ 时, 应该拒绝没有交互作用的假设.

例5.2 一种化学产品中的杂质受两个因子 (即压力与温度) 的影响. 由析因实验的单次重复所得的数据如表 5.10 所示. 平方和为

$$SS_A=\frac{1}{b}\sum_{i=1}^{a}y_{i.}^2-\frac{y_{..}^2}{ab}=\frac{1}{5}[23^2+13^2+8^2]-\frac{44^2}{(3)(5)}=23.33$$

$$SS_B=\frac{1}{a}\sum_{j=1}^{b}y_{.j}^2-\frac{y_{..}^2}{ab}=\frac{1}{3}[9^2+6^2+13^2+6^2+10^2]-\frac{44^2}{(3)(5)}=11.60$$

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{ab} = 166 - 129.07 = 36.93$$

$$SS_{残差} = SS_T - SS_A - SS_B = 36.93 - 23.33 - 11.60 = 2.00$$

由 (5.23) 式算得的非可加性部分的平方和为

$$\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}y_{i.}y_{.j} = (5)(23)(9) + (4)(23)(6) + \cdots + (2)(8)(10) = 7\ 236$$

$$SS_N = \frac{\left[\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}y_{i.}y_{.j} - y_{..}\left(SS_A + SS_B + \frac{y_{..}^2}{ab}\right)\right]^2}{abSS_ASS_B}$$

$$= \frac{[7\ 236 - (44)(23.33 + 11.60 + 129.07)]^2}{(3)(5)(23.33)(11.60)} = \frac{[20.00]^2}{4\ 059.42} = 0.098\ 5$$

由 (5.24) 式, 误差平方和是

$$SS_{误差} = SS_{残差} - SS_N = 2.00 - 0.098\ 5 = 1.901\ 5$$

表 5.10 例 5.2 的杂质数据

温度 (°F)	压力					
	25	30	35	40	45	$y_{i.}$
100	5	4	6	3	5	23
125	3	1	4	2	3	13
150	1	1	3	1	2	8
$y_{.j}$	9	6	13	6	10	$44 = y_{..}$

完整的方差分析概括在表 5.11 中, 非可加性部分的检验统计量是 $F_0 = 0.098\ 5/0.271\ 6 = 0.36$, 所以结论是, 在这些数据中, 没有证据表明数据中存在交互作用. 温度和压力的主效应是显著的.

表 5.11 例 5.2 的方差分析表

方差来源	平方和	自由度	均方	F_0	P 值
温度	23.33	2	11.67	42.97	0.000 1
压力	11.60	4	2.90	10.68	0.004 2
非可加性	0.098 5	1	0.098 5	0.36	0.567 4
误差	1.901 5	7	0.271 6		
总和	36.93	14			

在结束本节时, 我们注意到, 每单元只有一个观测值的二因子析因设计模型 [(5.22) 式] 看起来很像随机化完全区组设计模型 [(4.1) 式]. 实际上 Tukey 关于非可加性的单自由度的检验法可以直接用来检验随机化区组模型的交互作用. 然而, 导出随机化区组模型和析因模型的实验情况是十分不同的. 在析因模型中, 所有 ab 个试验都是以随机顺序进行的, 而在随机化区组模型中, 随机化仅在区组之内. 区组是一个随机化约束. 因而, 收集数据的方法以及对两个模型的解释都是完全不同的.

5.4 一般的析因设计

二因子析因设计的结果可以推广至一般情况, 其中因子 A 有 a 个水平, 因子 B 有 b 个水平, 因子 C 有 c 个水平, 等等, 这些因子安排在一个析因实验中. 一般说来, 当有完全实验的 n

次重复时, 将有 $abc\cdots n$ 个总的观测值. 当所有可能的交互作用包括在模型之内时, 为了确定误差平方和, 必须至少进行两次重复 ($n \geqslant 2$).

如果实验的所有因子都是固定的, 则容易给出计算公式并检验关于主效应和交互作用的假设. 对固定效应模型说来, 主效应和交互作用的检验统计量都可以将相应的主效应或交互作用的均方除以误差均方而得. 所有这些 F 检验都是上尾部的、单边的检验. 任一主效应的自由度是因子的水平数减 1, 交互作用的自由度是和交互作用有关的各个成分的自由度的乘积.

例如, 考虑**三因子方差分析模型**:

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk}$$

$$+(\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl} \quad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b \\ k = 1, 2, \cdots, c \\ l = 1, 2, \cdots, n \end{cases} \tag{5.26}$$

设 A, B, C 是固定的, **方差分析表**如表 5.12 所示. 关于主效应和交互作用的 F 检验可直接由期望均方得出.

表 5.12 三因子固定效应模型的方差分析

方差来源	平方和	自由度	均　方	期望均方	F_0
A	SS_A	$a-1$	MS_A	$\sigma + \frac{bcn\sum\tau_i^2}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
B	SS_B	$b-1$	MS_B	$\sigma^2 + \frac{acn\sum\beta_j^2}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
C	SS_C	$c-1$	MS_C	$\sigma^2 + \frac{abn\sum\gamma_k^2}{c-1}$	$F_0 = \frac{MS_C}{MS_E}$
AB	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	$\sigma^2 + \frac{cn\sum\sum(\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
AC	SS_{AC}	$(a-1)(c-1)$	MS_{AC}	$\sigma^2 + \frac{bn\sum\sum(\tau\gamma)_{ik}^2}{(a-1)(c-1)}$	$F_0 = \frac{MS_{AC}}{MS_E}$
BC	SS_{BC}	$(b-1)(c-1)$	MS_{BC}	$\sigma^2 + \frac{an\sum\sum(\beta\gamma)_{jk}^2}{(b-1)(c-1)}$	$F_0 = \frac{MS_{BC}}{MS_E}$
ABC	SS_{ABC}	$(a-1)(b-1)(c-1)$	MS_{ABC}	$\sigma^2 + \frac{n\sum\sum\sum(\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$	$F_0 = \frac{MS_{ABC}}{MS_E}$
误差	SS_E	$abc(n-1)$	MS_E	σ^2	
总和	SS_T	$abcn-1$			

通常, 利用统计软件包计算方差分析, 然而, 偶尔手工计算平方和的公式也是有用的. 总平方和由通常的办法得出:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{abcn} \tag{5.27}$$

主效应的平方和由因子 A, B, C 的总和 $y_{i...}$, $y_{.j..}$, $y_{..k.}$ 求得如下.

$$SS_A = \frac{1}{bcn}\sum_{i=1}^{a} y_{i...}^2 - \frac{y_{....}^2}{abcn} \tag{5.28}$$

$$SS_B = \frac{1}{acn}\sum_{j=1}^{b} y_{.j..}^2 - \frac{y_{....}^2}{abcn} \tag{5.29}$$

$$SS_C = \frac{1}{abn}\sum_{k=1}^{c} y_{..k.}^2 - \frac{y_{....}^2}{abcn} \tag{5.30}$$

计算二因子交互作用的平方和, 要用到 $A\times B, A\times C$, $B\times C$ 单元的总和. 要计算这些量, 通常借助于把原始数据表压缩为三个双向表. 由下式得出这些平方和:

$$SS_{AB} = \frac{1}{cn}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij..}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_B = SS_{\text{小计}(AB)} - SS_A - SS_B \tag{5.31}$$

$$SS_{AC} = \frac{1}{bn}\sum_{i=1}^{a}\sum_{k=1}^{c} y_{i.k.}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_C = SS_{\text{小计}(AC)} - SS_A - SS_C \tag{5.32}$$

$$SS_{BC} = \frac{1}{an}\sum_{j=1}^{b}\sum_{k=1}^{c} y_{.jk.}^2 - \frac{y_{....}^2}{abcn} - SS_B - SS_C = SS_{\text{小计}(BC)} - SS_B - SS_C \tag{5.33}$$

二因子小计的平方和从每个双向表的总和中得出. 三因子交互作用的平方和由三项单元总和 $\{y_{ijk.}\}$ 算得为

$$SS_{ABC} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} y_{ijk.}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} \tag{5.34a}$$

$$= SS_{\text{小计}(ABC)} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} \tag{5.34b}$$

误差平方和可以从总平方和减去每个主效应以及交互作用的平方和而得到, 或者用

$$SS_E = SS_T - SS_{\text{小计}(ABC)} \tag{5.35}$$

例 5.3 软饮料的灌装问题

一位软饮料灌装员感兴趣的问题是, 在他的生产线上生产的瓶装饮料有更为一致的灌装高度. 灌注机械从理论上说能以正确的标准高度来灌注每只饮料瓶, 但实际上, 围绕这一目标是有偏差的, 灌装员希望更好地了解这种变异性的来源进而减小这种偏差.

在进行灌注时, 生产线上的工程师能够控制 3 个变量: 碳酸百分率 (A), 灌注器的操作压强 (B), 每分钟生产的瓶数或流水线速度 (C). 压强和速度易于控制, 但是在实际生产时, 碳酸百分率却较难于控制, 因为它随产品温度的变化而变化. 不过, 为了实验, 工程师可以在 3 个水平上控制碳酸: 10%, 12%以及 14%. 他选取两种压强水平 (20 psi 与 30 psi) 和两种流水线速度水平 (200 bpm 与 250 bpm). 他决定在这三因子的析因设计中做两次重复, 所有 24 个试验的顺序是随机选取的. 观测到的响应变量是相对于目标灌注高度的平均偏差, 它是在每一组条件下生产瓶装饮料时观测到的. 从这一实验中所得的数据如表 5.13 所示. 正偏差表示灌注高度在目标之上, 负偏差表示灌注高度在目标之下. 表 5.13 中圆圈内的数是三向单元总和 $y_{ijk.}$.

由 (5.27) 式得出总校正平方和为

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{abcn} = 571 - \frac{(75)^2}{24} = 336.625$$

由 (5.28) 式、(5.29) 式与 (5.30) 式算得主效应的平方和是

$$SS_{\text{碳酸}} = \frac{1}{bcn}\sum_{i=1}^{a} y_{i...}^2 - \frac{y_{....}^2}{abcn} = \frac{1}{8}[(-4)^2 + (20)^2 + (59)^2] - \frac{(75)^2}{24} = 252.750$$

$$SS_{\text{压强}} = \frac{1}{acn}\sum_{j=1}^{b} y_{.j..}^2 - \frac{y_{....}^2}{abcn} = \frac{1}{12}[(21)^2 + (54)^2] - \frac{(75)^2}{24} = 45.375$$

$$SS_{\text{速度}} = \frac{1}{abn}\sum_{k=1}^{c} y_{..k.}^2 - \frac{y_{....}^2}{abcn} = \frac{1}{12}[(26)^2 + (49)^2] - \frac{(75)^2}{24} = 22.042$$

表 5.13 例 5.3 的罐装高度偏差数据

碳酸百分率(A)	操作压强 (B) 25 psi 流水线速度(C) 200	25 psi 250	30 psi 流水线速度(C) 200	30 psi 250	$y_{i...}$
10	−3, −1 (−4)	−1, 0 (−1)	−1, 0 (−1)	1, 1 (2)	−4
12	0, 1 (1)	2, 1 (3)	2, 3 (5)	6, 5 (11)	20
14	5, 4 (9)	7, 6 (13)	7, 9 (16)	10, 11 (21)	59
$B\times C$ 总和 $y_{.jk.}$	6	15	20	34	$75 = y_{....}$
$y_{.j..}$	21		54		

$A\times B$ 总和 $y_{ij..}$

A	B 25	30
10	−5	1
12	4	16
14	22	37

$A\times C$ 总和 $y_{i.k.}$

A	C 200	250
10	−5	1
12	6	14
14	25	34

要计算二因子交互作用的平方和, 必须求出双向单元的总和. 例如, 为了求碳酸–压强即 AB 的交互作用, 需要表 5.13 中 $A\times B$ 的单元总和 $\{y_{ij..}\}$. 用 (5.31) 式, 求得平方和为

$$SS_{AB} = \frac{1}{cn}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij..}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_B$$

$$= \frac{1}{4}[(-5)^2 + (1)^2 + (4)^2 + (16)^2 + (22)^2 + (37)^2] - \frac{(75)^2}{24}$$

$$- 252.750 - 45.375 = 5.250$$

碳酸–速度即 AC 的交互作用由表 5.13 所示的 $A\times C$ 的单元总和 $\{y_{i.k.}\}$ 以及 (5.32) 式求得

$$SS_{AC} = \frac{1}{bn}\sum_{i=1}^{a}\sum_{k=1}^{c} y_{i.k.}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_C$$

$$= \frac{1}{4}[(-5)^2 + (1)^2 + (6)^2 + (14)^2 + (25)^2 + (34)^2] - \frac{(75)^2}{24}$$

$$- 252.750 - 22.042 = 0.583$$

压强–速度即 BC 的交互作用由表 5.13 的 $B\times C$ 的单元总和 $\{y_{.jk.}\}$ 以及 (5.33) 式求得:

$$SS_{BC} = \frac{1}{an}\sum_{j=1}^{b}\sum_{k=1}^{c} y_{.jk.}^2 - \frac{y_{....}^2}{abcn} - SS_B - SS_C$$

$$= \frac{1}{6}[(6)^2 + (15)^2 + (20)^2 + (34)^2] - \frac{(75)^2}{24} - 45.375 - 22.042 = 1.042$$

三因子交互作用的平方和由 $A\times B\times C$ 的单元总和 $\{y_{ijk.}\}$ 求得, 那是表 5.13 中用圆圈圈起来的数. 由 (5.34a) 式, 得

$$\begin{aligned}SS_{ABC} &= \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} y_{ijk.}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}\\ &= \frac{1}{2}[(-4)^2 + (-1)^2 + (-1)^2 + \cdots + (16)^2 + (21)^2] - \frac{(75)^2}{24}\\ &\quad - 252.750 - 45.375 - 22.042 - 5.250 - 0.583 - 1.042 = 1.083\end{aligned}$$

最后, 注意到

$$SS_{\text{小计}(ABC)} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} y_{ijk.}^2 - \frac{y_{....}^2}{abcn} = 328.125$$

我们有

$$SS_E = SS_T - SS_{\text{小计}(ABC)} = 336.625 - 328.125 = 8.500$$

方差分析概括在表 5.14 中. 碳酸百分率、操作压强、流水线速度都显著影响灌注量. 碳酸–压强交互作用的 F 比的 p 值为 0.055 8, 这表明这两个因子之间有弱的交互作用.

表 5.14 例 5.3 的方差分析表

方差来源	平方和	自由度	均　方	F_0	P 值
碳酸百分率 (A)	252.750	2	126.375	178.412	< 0.000 1
操作压强 (B)	45.375	1	45.375	64.059	< 0.000 1
流水线速度 (C)	22.042	1	22.042	31.118	0.000 1
AB	5.250	2	2.625	3.706	0.055 8
AC	0.583	2	0.292	0.412	0.671 3
BC	1.042	1	1.042	1.471	0.248 5
ABC	1.083	2	0.542	0.765	0.486 7
误差	8.500	12	0.708		
总和	336.625	23			

下一步是这一实验的残差分析. 这一点留给读者作为练习. 值得指出的是, 残差的正态概率图以及其他通常的诊断都没有显示出任何异常.

为了从实用上进一步说明这一实验, 图 5.16 给出 3 个主效应和 AB(碳酸–压强) 交互作用的图形. 主效应图形刻画了在 3 个因子的水平处的边际响应平均值. 所有 3 个变量都有正的主效应, 也就是说, 当变量增大时, 相对于灌注目标的平均偏差变大. 碳酸与压强之间的交互作用相当小, 如图 5.16d 所示, 两条曲线的形状相似.

因为公司要求相对于灌注目标的平均偏差接近于零, 工程师决定推荐低水平的操作压强 (25 psi) 和高水平的流水线速度 (250 bpm, 它将使生产率最大化). 图 5.17 刻画了在这组操作条件下, 在 3 种不同的碳酸水平上观测到的相对于目标灌注高度的平均偏差. 现在, 碳酸水平在生产过程中不能被完全控制, 图 5.17 中的实曲线表示的正态分布近似地表示了当前实验中碳酸水平的变化. 当生产过程受这一分布所刻画的碳酸水平的值所影响时, 灌注高度有明显的升降. 如果碳酸水平值的分布服从如图 5.17 虚线所示的正态分布, 则灌注高度的变化会减小. 在生产时改善温度控制就能减小碳酸水平分布的标准差.

我们曾经指出过, 如果析因实验中所有的因子都是固定的, 则可以直接构造检验统计量. 检验任一主效应或交互作用的统计量通常都由主效应或交互作用的均方除以误差均方而得. 不过, 当析因实验涉及一个或多个**随机因子**时, 检验统计量的构造就不是这样的了. 我们需要考察均方的期望才能确定正确的检验法. 如此看来, 我们需要有一种导出均方期望值的方法. 我们将在第 12 章全面讨论这一课题.

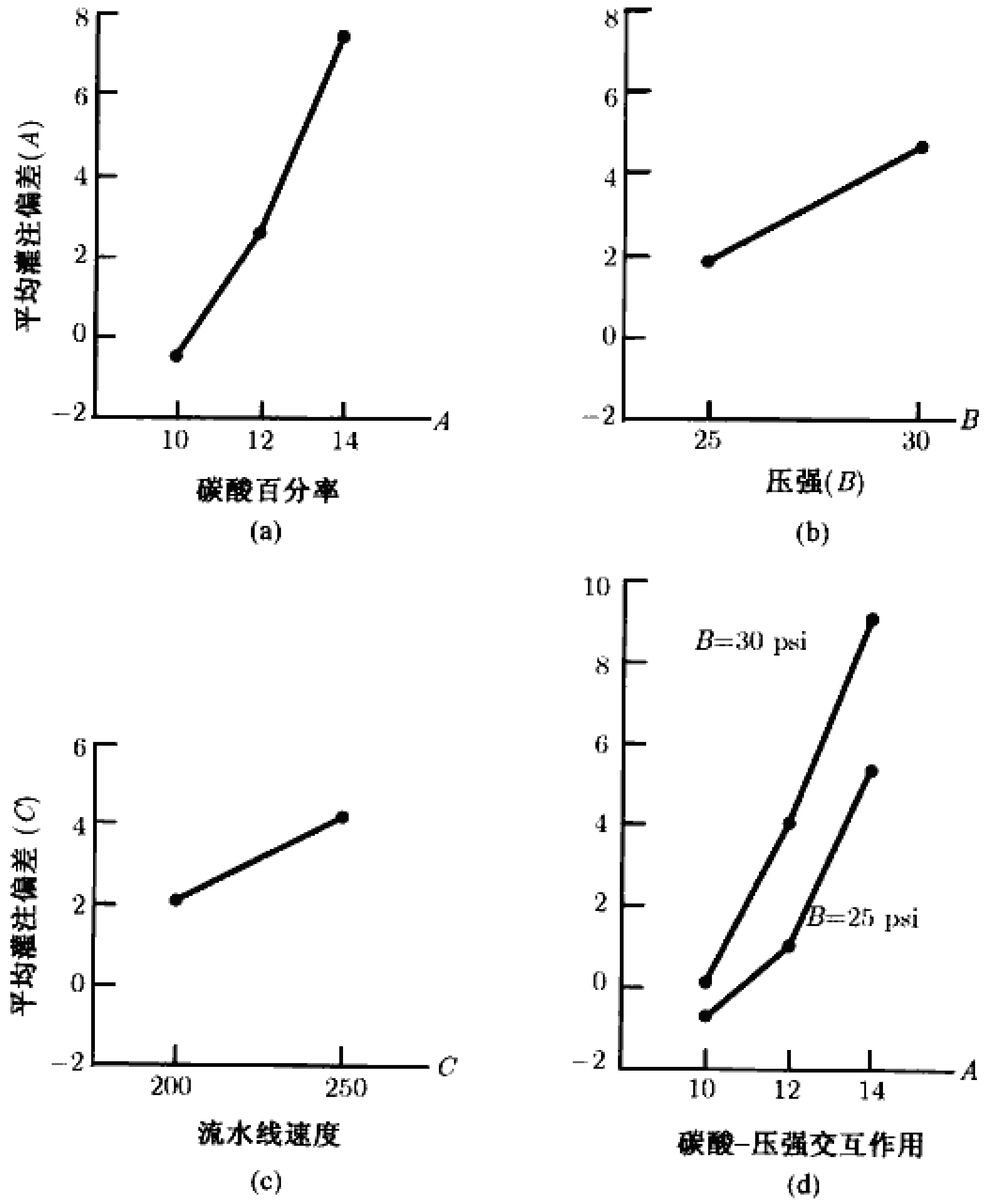


图 5.16 例 5.3 的主效应和交互作用图

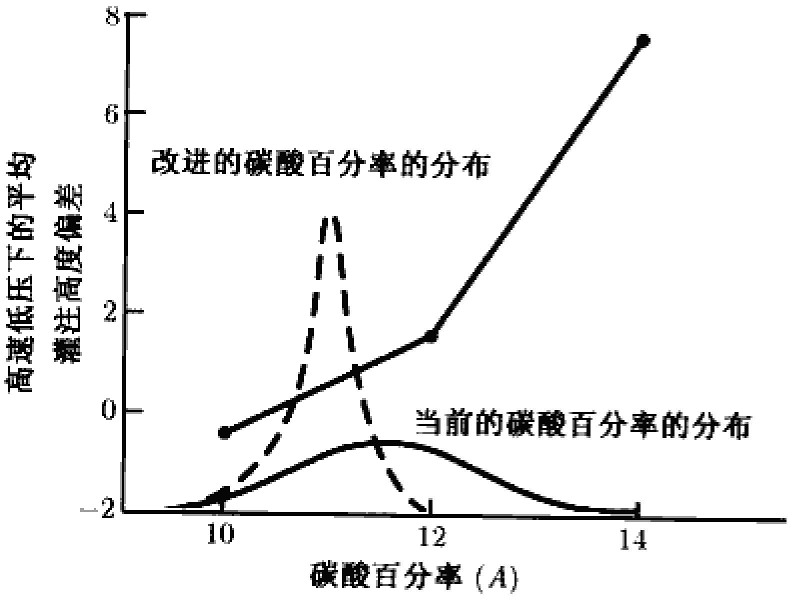


图 5.17 在高速低压下不同碳酸水平的平均灌注高度偏差

5.5 拟合响应曲线与曲面

我们注意到, 给一定量因子的水平拟合一条响应曲线(response curve) 是有用的, 它使实验

者可以得到一个响应与因子的关系式. 这一关系式可以用来进行插值, 也就是说, 用来预测在因子水平上的响应, 当然, 这些水平应处在实验中实际使用过的水平之间. 当至少有两个因子是定量时, 我们可以在这些设计因子的不同组合上拟合预测值 y 的**响应曲面**(response surface). 一般地, **线性回归方法**用于为这些实验数据拟合相应的模型. 我们在 3.5.1 节中对于单因子的实验说明了这个方法, 并列举了涉及析因实验的两个例子. 在这些例子中, 我们用计算机软件包生成了回归模型. 关于回归分析的更多信息, 见第 10 章以及本章的补充材料.

例 5.4 考虑例 5.1 中的实验. 因子温度是定量的, 材料类型是定性的. 且温度有 3 个水平. 因此, 我们可计算线性的和二次的温度效应, 以便研究温度怎样影响电池寿命. 表 5.15 给出了该实验的 Design - Expert 主要输出, 其中假定温度是定量的, 材料类型是定性的.

表 5.15 中的 ANOVA 表明 "模型" 来源的变异性已分解为几个分量. 分量 "A" "A^2" 分别表示温度的一次效应、二次效应, "B" 表示材料类型因子的主效应. 材料类型是具有 3 个水平的定性因子. 项 "AB" 和 "A^2B" 分别是线性温度因子及二次的温度因子和材料类型的交互作用.

P 值表明 A^2 和 AB 不显著, 而 A^2B 是显著的. 通常我们考虑从模型中移去不显著的项或因子, 但是, 此时移去 A^2 和 AB, 保留 A^2B 导致模型是不分层的. **分层原理**(hierarchy principle) 表明如果一个模型包含高阶项 (比如 A^2B), 它必须包含分解它的所有低阶的项 (这里指 A^2 和 AB). 分层提升了模型的内部一致性, 许多统计模型创建者严格遵从这个原理. 但是, 分层并不总是一个好主意, 预测公式中不含促成分层的不显著项的模型确实工作得更好. 更多信息, 见本章的补充材料.

计算机输出也给出了模型的系数估计, 以及规范因子 (coded factor) 的电池寿命的最终预测公式. 该式中, 当温度为低, 中, 高水平 (15°, 70°, 125°), 对应的温度水平分别为 $A = -1, 0, +1$. 变量 $B[1]$ 和 $B[2]$ 被规范为**示性变量**(indicator variable), 其定义为:

	材料类型		
	1	2	3
$B[1]$	1	0	−1
$B[2]$	0	1	−1

也有用因子实际水平预测电池寿命的公式. 因为材料类型是定性因子, 所以对于每一种材料类型, 有一个作为温度的函数的寿命预测公式. 图 5.18 显示了由这 3 个预测公式生成的响应曲面. 将它们与图 5.9 中的该实验的二因子交互作用图进行比较.

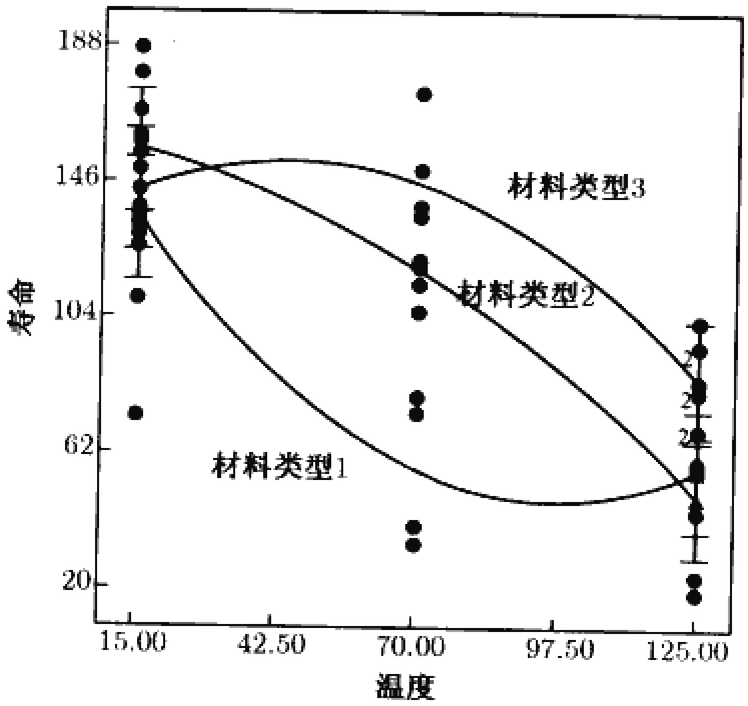


图 5.18 例 5.4 中, 对于这 3 种材料类型, 作为温度函数的预测寿命

表 5.15 例 5.4 的 Design-Expert 输出

Response:Life in hr

ANOVA for Response Surface Reduced Cubic Model

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	*F* Value	Prov> *F*	
Model	59416.22	8	7427.03	11.00	< 0.0001	significant
A	39042.67	1	39042.67	57.82	< 0.0001	
B	10683.72	2	5341.86	7.91	0.0020	
A^2	76.06	1	76.06	0.11	0.7398	
AB	2315.08	2	1157.54	1.71	0.1991	
A^2B	7298.69	2	3649.35	5.40	0.0106	
Residual	18230.75	27	675.21			
Lack of Fit	0.000	0				
Pure Error	18230.75	27	675.21			
Cor Total	77646.97	35				

Std.Dev.	25.98	R-Squared	0.7652
Mean	105.53	Adj R-Squared	0.6956
C.V.	24.62	Pred R-Squared	0.5826
PRESS	32410.22	Adeq Precision	8.178

Term	Coefficient Estimate	DF	Standard Error	95%CI Low	95%CI High	VIF
Intercept	107.58	1	7.50	92.19	122.97	
A-Temp	−40.33	1	5.30	−51.22	−29.45	1.00
B[1]	−50.33	1	10.61	−72.10	−28.57	
B[2]	12.17	1	10.61	−9.60	33.93	
A^2	−3.08	1	9.19	−21.93	15.77	1.00
AB[1]	1.71	1	7.50	−13.68	17.10	
AB[2]	−12.79	1	7.50	−28.18	2.60	
A^2B[1]	41.96	1	12.99	15.30	68.62	
A^2B[2]	−14.04	1	12.99	−40.70	12.62	

Final Equation in Terms of Coded Factors:

Life=
+107.58
−40.33 *A
−50.33 *B[1]
+12.17 *B[2]
−3.08 *A^2
+1.71 *AB[1]
−12.79 *AB[2]
+41.96 *A^2B[1]
−14.04 *A^2[2]

Final Equation in Terms of Actual Factors:

Material type 1
Life=
+ 169.38017
−2.48860 *Temp
+0.012851 *Temp^2

Material type 2
Life=
+159.62397
−0.17901 *Temp
+0.41627 *Temp^2

Material Type 3
Life=
+132.76240
+0.89264 *Temp
−0.43218 *Temp^2

如果析因实验中的几个因子是定量的，**响应曲面**也许可用于模拟 y 和设计因子的关系，而且这些定量因子效应可表示为单自由度的多项式效应. 类似地，可将定量因子的交互作用分解为单自由度的交互作用分量. 下述例子说明这一点.

例 5.5 人们猜想，装配在一台数控机床上的切割工具的有效寿命受切割速度和工具角度的影响. 选用 3 种速度和 3 种角度，实施两次重复的 3^2 析因实验，规范数据如表 5.16 所示. 单元中圈起来的数是单元的总和 $\{y_{ij.}\}$.

表 5.16 工具寿命实验的数据

工具角度(度)	切割速度 (英寸/分钟)			
	125	150	175	$y_{i..}$
15	−2, −1 (−3)	−3, 0 (−3)	2, 3 (5)	−1
20	0, 2 (2)	1, 3 (4)	4, 6 (10)	16
25	−1, 0 (−1)	5, 6 (11)	0, −1 (−1)	9
$y_{.j.}$	−2	12	14	$24=y_{...}$

表 5.17 给出了该例的 Minitab 输出. 表的上半部包含标准 ANOVA，两个设计因子都视为定性因子. 然而本例中的两个因子都是定量的，表的中间部分的回归分析用到了这一点. 项 a 和项 $a2$ 是工具角度的线性效应和二次效应，b 和 $b2$ 是速度的线性效应和二次效应. 项 ab, $a2b$, $ab2$, $a2b2$ 代表了二因子交互作用的线性 × 线性、二次 × 线性、线性 × 二次、二次 × 二次分量. 尽管有一些大的 P 值，我们仍保留所有模型项以确保分层. 用实际因子单位来表示预测公式. 表的底部列出了每个模型项的平方和. (它们是顺序平方和，缩写为 "Seq SS"). 注意到，$SS_{\text{工具角度}} = SS_a + SS_{a2}, SS_{\text{切割速度}} = SS_b + SS_{b2}$, $SS_{\text{角度}\times\text{速度}} = SS_{ab} + SS_{a2b} + SS_{ab2} + SS_{a2b2}$，因为 3^2 析因设计是**正交设计**，所以这些顺序平方和完全分解了模型平方和.

图 5.19 给出了由工具寿命的预测公式生成的曲面的**等高线图**. 检查这个**响应曲面**，从中可以看出工具寿命在切割速度为 150 转/分和角度为 25° 时达到极大值. 图 5.20 的三维响应曲面图给出实质上相同的信息，但它给出了一个不同的、有时会更有用的关于工具寿命响应曲面的透视图. 研究响应曲面是实验设计的一个重要方面，第 11 章中将详细讨论它.

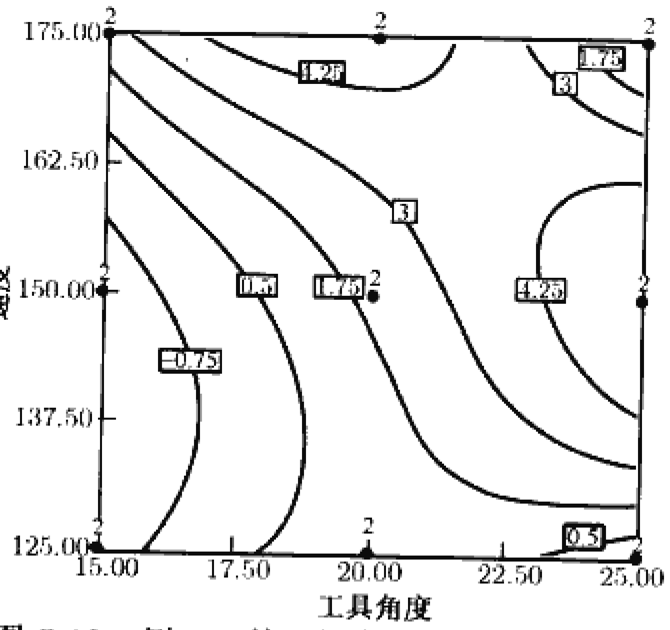


图 5.19 例 5.5 的工具寿命的预测等式生成的曲面的等高线图

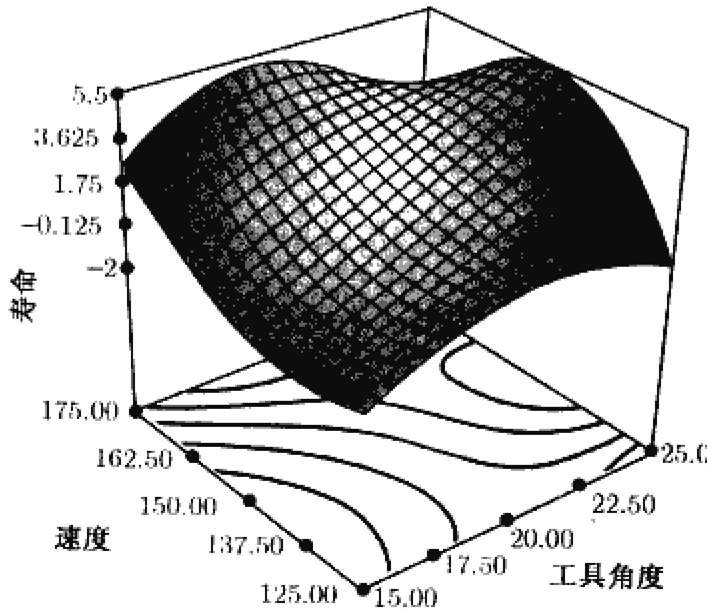


图 5.20 例 5.5 的工具寿命的三维响应曲面图

表 5.17 例 5.5 的 Minitab 输出

```
ANOVA:Tool Life versus Tool Angle, Cutting Speed

Factor          Type      Levels    Values
Tool Angle      fixed          3    15,20,25
Cutting Speed   fixed          3    125,150,175

Analysis of Variance for Tool Life

Source                        DF        SS        MS        F        P
Tool Angle                     2    24.333    12.167     8.42    0.009
Cutting Speed                  2    25.333    12.667     8.77    0.008
Tool Angle* Cutting Speed      4    61.333    15.333    10.62    0.002
Error                          9    13.000     1.444
Total                         17   124.000

S=1.20185        R-Sq=89.52%        R-Sq(adj)=80.20%

Regression Analysis: Tool  Life  versus a, b, a2, b2, ab, ab2, a2b, a2b2

The regression equation is
Tool life = - 1068 + 136 a + 14.5 b - 4.08 a2 - 0.0496 b2 - 1.86 al
          + 0.00640 ab2+0.0560 a2b - 0.000192 a2b2

Predictor          Coef       SE Coef        T        P
Constant        -1068.0         702.2    -1.52    0.163
a                136.30         72.61     1.88    0.093
b                14.480         9.503     1.52    0.162
a2               -4.080         1.810    -2.25    0.051
b2             -0.04960       0.03164    -1.57    0.151
ab              -1.8640        0.9827    -1.90    0.090
ab2            0.006400      0.003272     1.96    0.082
a2b             0.05600       0.02450     2.29    0.048
a2b2        -0.00019200    0.00008158    -2.35    0.043

S=1.20185        R-Sq=89.5%        R-Sq(adj)=80.2%

Analysis of Variance

Source            DF        SS        MS       F       P
Regression         8   111.000    13.875    9.61   0.001
Residual Error     9    13.000     1.444
Total             17   124.000

Source   DF     SS
a         1    8.333  ──┐
b         1   21.333  ──┼──> SS(Tool Angle)=8.333+16.000=24.333
a2        1   16.000  ──┘
b2        1    4.000  ─────> SS(Cutting Speed)=21.333+4.000=25.333
ab        1    8.000  ┐
ab2       1   42.667  ├────> SS(Angle×Speed)=8.000+42.667+2.667+8.000=61.33
a2b       1    2.667  │
a2b2      1    8.000  ┘
```

5.6 析因设计中的区组化

我们已经讨论了**完全随机化**实验下的析因设计. 在析因设计中完全随机化所有试验有时是不现实的. 例如, 讨厌因子的存在也许要求以**区组**形式进行实验. 我们在第 4 章中讨论了单因子实验的区组的基本概念, 现在讨论如何在析因设计中引进区组. 析因设计区组化的其他内容将在

第 7, 8, 9 章和第 13 章中介绍.

考虑具有二因子 (A 和 B) 及 n 次重复的析因实验. 该设计的线性统计模型为

$$y_{ijk}=\mu+\tau_i+\beta_j+(\tau\beta)_{ij}+\varepsilon_{ijk}\quad\begin{cases}i=1,2,\cdots,a\\j=1,2,\cdots,b\\k=1,2,\cdots,n\end{cases}\tag{5.36}$$

其中 $\tau_i,\beta_j,(\tau\beta)_{ij}$ 分别表示因子 A,B, AB 交互作用. 假定进行这个实验需要特殊原材料. 这种原材料有许多批次, 但是每批都没有大到足以在同一批次上进行所有 abn 个处理组合. 但是, 如果一个批次包含的材料足够 ab 次观测, 则实验设计也可以用单独批次的原材料进行 n 次重复中的每一次. 于是, 原材料的批次表示了随机化约束或区组, 在每个区组内进行完全析因设计的一次重复. 新设计的效应模型为

$$y_{ijk}=\mu+\tau_i+\beta_j+(\tau\beta)_{ij}+\delta_k+\epsilon_{ijk}\quad\begin{cases}i=1,2,\cdots,a\\j=1,2,\cdots,b\\k=1,2,\cdots,n\end{cases}\tag{5.37}$$

其中 δ_k 是第 k 个区组的效应. 当然, 区组内的处理组合的进行次序是完全随机化的.

模型 [(5.37) 式] 假定区组和处理间的交互作用是可以忽略的. 这在以前随机化区组设计中曾假设过. 如果交互作用确实存在, 则它们不能从误差分量中分离出来. 事实上, 这个模型的误差项包含了 $(\tau\delta)_{ik},(\beta\delta)_{jk},(\tau\beta\delta)_{ijk}$ 的交互作用. 方差分析列在表 5.18 中. 该表与析因设计类似, 其误差平方和加上区组平方和等于析因设计的误差平方和. 我们用 n 个区组总和$\{y_{..k}\}$间的平方和来计算区组平方和.

表 5.18　随机化完全区组中的二因子析因设计的方差

方差来源	平方和	自由度	期望均方	F_0
区组	$\frac{1}{ab}\sum\limits_k y_{..k}^2-\frac{y_{...}^2}{abn}$	$n-1$	$\sigma^2+ab\sigma_\delta^2$	
A	$\frac{1}{bn}\sum\limits_i y_{i..}^2-\frac{y_{...}^2}{abn}$	$a-1$	$\sigma^2+\frac{bn\sum\tau_i^2}{a-1}$	$\frac{MS_A}{MS_E}$
B	$\frac{1}{an}\sum\limits_j y_{.j.}^2-\frac{y_{...}^2}{abn}$	$b-1$	$\sigma^2+\frac{an\sum\beta_j^2}{b-1}$	$\frac{MS_B}{MS_E}$
AB	$\frac{1}{n}\sum\limits_i\sum\limits_j y_{ij.}^2-\frac{y_{...}^2}{abn}-SS_A-SS_B$	$(a-1)(b-1)$	$\sigma^2+\frac{n\sum\sum(\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_E}$
误差	减法	$(ab-1)(n-1)$	σ^2	
总和	$\sum\limits_i\sum\limits_j\sum\limits_k y_{ijk}^2-\frac{y_{...}^2}{abn}$	$abn-1$		

在前面的例子中, 随机化被限制在原材料的同一批次中. 实际上, 多种现象也可以产生随机化约束, 如时间和操作员. 例如, 我们不能在一天内进行完整的析因实验, 实验者可以在第一天做一次完全重复, 在第二天做一次完全重复, 依此类推. 结果, 每一天将作为一个区组.

例 5.6　工程师研究提高雷达示波器探测目标能力的方法. 她认为有两个因子是重要的: 一是示波器内的背景噪音量 (即 "地面杂乱回波"), 二是显示器的滤波器的型号. 实验者计划用 3 种水平的地面杂乱回波和两种型号的滤波器. 可以认为它们是固定类型的因子. 随机选择处理组合 (地面杂乱回波水平和滤波器

型号) 进行实验并引入表示进入示波器内的目标的信号. 目标的亮度逐渐增加直到操作员能观测到它. 作为响应变量, 所检验的亮度水平是可以度量的. 因为操作员足够多, 所以总能选择一位操作员, 使得他或她在示波器上完成要求的试验. 又因为操作员使用示波器的技巧和能力有所不同. 因此, 可以将操作员作为区组. 随机选择 4 位操作员. 一旦操作员被选定, 6 个处理组合进行的次序也被随机确定. 因此, 我们在完全随机化区组中进行 3×2 析因实验. 数据显示在表 5.19 中.

该实验的线性模型为

$$y_{ijk}=\mu+\tau_i+\beta_j+(\tau\beta)_{ij}+\delta_k+\varepsilon_{ijk}\quad\begin{cases}i=1,2,3\\ j=1,2\\ k=1,2,3,4\end{cases}$$

其中 τ_i 表示地面杂乱回波效应, β_j 表示滤波器型号效应, $(\tau\beta)_{ij}$ 表示交互作用, δ_k 是区组效应, ε_{ijk} 是 NID(0, σ^2) 误差分量. 地面杂乱回波、滤波器型号以及交互作用的平方和可以利用通常的方法计算. 区组平方和可以通过操作员总和计算如下:

$$SS_{区组}=\frac{1}{ab}\sum_{k=1}^{n}y_{..k}^2-\frac{y_{...}^2}{abn}=\frac{1}{(3)(2)}[(572)^2+(579)^2+(597)^2+(530)^2]-\frac{(2\ 278)^2}{(3)(2)(4)}=402.17$$

表 5.19 目标探测处的亮度水平

操作员 (区组)	1		2		3		4	
滤波器型号	1	2	1	2	1	2	1	2
地物干扰								
低	90	86	96	84	100	92	92	81
中	102	87	106	90	105	97	96	80
高	114	93	112	91	108	95	98	83

该实验的完整方差分析概括在表 5.20 中. 表 5.20 通过所有效应的均方和除以均方误差来检验它们. 在 1%水平上, 地面杂乱回波和滤波器型号都是显著的, 而交互作用仅在 10%的水平上显著. 因此我们得出结论: 所用的地面杂乱回波和滤波器型号都影响了操作员探测目标的能力. 也有证据表明这些因子间存在着适度的交互作用.

表 5.20 例 5.6 的方差分析表

方差来源	平方和	自由度	均方	F_0	P 值
地面杂乱回波 (G)	335.58	2	167.79	15.13	0.000 3
滤波器型号 (F)	1 066.67	1	1 066.67	96.19	<0.000 1
GF	77.08	2	38.54	3.48	0.057 3
区组	402.17	3	134.06		
误差	166.33	15	11.09		
总和	2 047.83	23			

在有两个随机化约束, 每个约束有 p 个水平的情形下, 如果在 k 因子析因设计中处理组合的数量精确等于约束水平数, 即 $p=ab\cdots m$, 则析因设计可以在 $p\times p$ 拉丁方中进行. 例如, 考虑修改过的例 5.6 的雷达目标探测实验. 实验的因子包括滤波器型号 (两个水平) 和地面杂乱回波 (3 个水平), 操作员视为区组. 现在假定对时间有要求, 每天只能做 6 次试验. 因而, 日期成为第二个随机约束, 结果为 6×6 拉丁方设计, 如表 5.21 所示. 在此表中, 我们用小写字母 f_i 和 g_j 分别表示滤波器型号的第 i 个水平和地面杂乱回波的第 j 个水平. 也就是, f_1g_2 代表滤波器型号 1 和中等的地面杂乱回波. 现在需要 6 名操作员, 而不是先前实验的 4 个, 因此在 3×2 析因设计中的处理组合数恰好等于约束水平数. 假定在该设计中, 每名操作员每天仅做一次试验.

拉丁字母 A, B, C, D, E, F 代表如下的 3×2 析因处理组合：$A = f_1g_1$, $B = f_1g_2$, $C = f_1g_3$, $D = f_2g_1$, $E = f_2g_2$, $F = f_2g_3$.

6 个拉丁字母间的 5 个自由度对应于滤波器型号 (1 个自由度)、地面杂乱回波 (2 个自由度) 以及它们的交互作用 (2 个自由度) 的主效应. 该设计的线性统计模型为

$$y_{ijkl} = \mu + \alpha_i + \tau_j + \beta_k + (\tau\beta)_{jk} + \theta_l + \varepsilon_{ijkl} \quad \begin{cases} i = 1, 2, \cdots, 6 \\ j = 1, 2, 3 \\ k = 1, 2 \\ l = 1, 2, \cdots, 6 \end{cases} \tag{5.38}$$

表 5.21 用 6×6 拉丁方进行的雷达探测实验

日期	操作员					
	1	2	3	4	5	6
1	$A(f_1g_1 = 90)$	$B(f_1g_2 = 106)$	$C(f_1g_3 = 108)$	$D(f_2g_1 = 81)$	$F(f_2g_3 = 90)$	$E(f_2g_2 = 88)$
2	$C(f_1g_3 = 114)$	$A(f_1g_1 = 96)$	$B(f_1g_2 = 105)$	$F(f_2g_3 = 83)$	$E(f_2g_2 = 86)$	$D(f_2g_1 = 84)$
3	$B(f_1g_2 = 102)$	$E(f_2g_2 = 90)$	$G(f_2g_3 = 95)$	$A(f_1g_1 = 92)$	$D(f_2g_1 = 85)$	$C(f_1g_3 = 104)$
4	$E(f_2g_2 = 87)$	$D(f_2g_1 = 84)$	$A(f_1g_1 = 100)$	$B(f_1g_2 = 96)$	$C(f_1g_3 = 110)$	$F(f_2g_3 = 91)$
5	$F(f_2g_3 = 93)$	$C(f_1g_3 = 112)$	$D(f_2g_1 = 92)$	$E(f_2g_2 = 80)$	$A(f_1g_1 = 90)$	$B(f_1g_2 = 98)$
6	$D(f_2g_1 = 86)$	$F(f_2g_3 = 91)$	$E(f_2g_2 = 97)$	$C(f_1g_3 = 98)$	$B(f_1g_2 = 100)$	$A(f_1g_1 = 92)$

其中 τ_j 和 β_k 分别表示地面杂乱回波和滤波器型号效应, α_i 和 θ_l 分别表示日期和操作员的随机化约束. 下面处理总和的双向表有助于计算平方和：

地面杂乱回波	滤波器型号 1	滤波器型号 2	$y_{\cdot j \cdot\cdot}$
低	560	512	1 072
中	607	528	1 135
高	646	543	1 189
$y_{\cdot\cdot k \cdot}$	1 813	1 583	3 396= $y_{\cdot\cdot\cdot\cdot}$

又, 行和与列和分别为

行 $(y_{\cdot jkl})$: 563 568 568 568 565 564

列 $(y_{ijk\cdot})$: 572 579 597 530 561 557

方差分析概括在表 5.22 中. 我们在表中增加了一列来表明每个平方和的自由度的数目是如何确定的.

表 5.22 作为拉丁方的 3×2 析因设计进行雷达探测实验的方差分析表

方差来源	平方和	自由度	自由度的一般公式	均方	F_0	P 值
地面杂乱回波,G	571.50	2	$a-1$	285.75	28.86	<0.000 1
滤波器型号,F	1 469.44	1	$b-1$	1 469.44	148.43	<0.000 1
GF	126.73	2	$(a-1)(b-1)$	63.37	6.40	0.007 1
日期 (行)	4.33	5	$ab-1$	0.87		
操作员 (列)	428.00	5	$ab-1$	85.60		
误差	198.00	20	$(ab-1)(ab-2)$	9.90		
总和	2 798.00	35	$(ab)^2-1$			

5.7 思 考 题

5.1 研究化学过程的产率. 设想两个最重要的变量是压强与温度. 每一因子选取 3 个水平, 进行有两次重复的析因实验. 产率数据如下:

温度 (℃)	压强 (psig)		
	200	215	230
150	90.4	90.7	90.2
	90.2	90.6	90.4
160	90.1	90.5	89.9
	90.3	90.6	90.1
170	90.5	90.8	90.4
	90.7	90.9	90.1

(a) 分析这些数据并写出结论. 用 $\alpha = 0.05$.

(b) 作出适当的残差图并论述模型的适合性.

(c) 你将在什么条件下运行这一生产过程?

5.2 一位工程师推测, 金属部件的表面光洁度受进料速度和切割深度的影响, 他选取 3 种进料速度并随机选取 4 种切割深度. 这样, 他实施了一个析因实验并得出下述数据:

进料速度 (in/min)	切割深度 (in)			
	0.15	0.18	0.20	0.25
	74	79	82	99
0.20	64	68	88	104
	60	73	92	96
	92	98	99	104
0.25	86	104	108	110
	88	88	95	99
	99	104	108	114
0.30	98	99	110	111
	102	95	99	107

(a) 分析这些数据并写出结论. 用 $\alpha = 0.05$.

(b) 作出适当的残差图并论述模型的适合性.

(c) 求在每种进料速度下平均表面光洁度的点估计.

(d) 确定 (a) 中检验的 P 值.

5.3 对思考题 5.2 中的数据, 对进料速度为 0.20 in/min 与 0.25 in/min 的响应均值差, 计算其 95%置信区间估计.

5.4 *Industrial Quality Control* (1956, pp.5~8) 上有一篇论文, 它描述了一个研究玻璃类型与荧光粉类型对电视显像管亮度的效应的实验. 响应变量是为得到规定亮度水平的电流必需量 (单位: 微安). 数据如下:

(a) 有何证据表明两种因子都影响亮度? 用 $\alpha = 0.05$.

(b) 两种因子有交互作用吗? 用 $\alpha = 0.05$.

(c) 分析这一实验的残差.

玻璃类型	荧光粉类型		
	1	2	3
1	280	300	290
	290	310	285
	285	295	290
2	230	260	220
	235	240	225
	240	235	230

5.5 Johnson 与 Leone (*Statistics and Experimental Design in Engineering and the Physical Science*, Wiley, 1977) 描述了一个研究铜板挠曲的实验. 所研究的两个因子是温度和铜板的含铜量. 响应变量是挠曲量的度量. 数据如下:

温度 (℃)	含铜量 (%)			
	40	60	80	100
50	17,20	16,21	24,22	28,27
75	12, 9	18,13	17,12	27,31
100	16,12	18,21	25,23	30,23
125	21,17	23,21	23,22	29,31

(a) 有证据表明两个因子都影响挠曲量吗? 因子之间有交互作用吗?

(b) 分析这一实验的残差.

(c) 作出各个含铜量水平的平均挠曲量的图像并与适当尺寸的 t 分布进行比较. 描述不同的含铜量水平对挠曲量的效应的差别. 如果要求低的挠曲量, 那么你会规定哪种含铜量水平?

(d) 设使用铜板的环境温度不易控制, 你会改变上述 (c) 的答案吗?

*5.6 研究影响合成纤维抗断强度的因子, 随机选取 4 台生产机器和 3 位操作员, 并用同一批产品的纤维进行析因实验. 结果如下:

操作员	机器			
	1	2	3	4
1	109	110	108	110
	110	115	109	108
2	110	110	111	114
	112	111	109	112
3	116	112	114	120
	114	115	119	117

(a) 分析这些数据并写出结论. 用 $\alpha = 0.05$.

(b) 作出恰当的残差图并论述模型的适合性.

5.7 一位机械工程师研究由钻头压力产生的推力. 他推测钻孔速度和材料的进料速度是两个最重要的因子. 随机选取 4 种进料速度, 并选用高和低的钻孔速度以代表极端的操作条件. 得出下述结果, 分析这些数据并写出结论. 用 $\alpha = 0.05$.

钻孔速度	进料速度			
	0.015	0.030	0.045	0.060
125	2.70	2.45	2.60	2.75
	2.78	2.49	2.72	2.86
200	2.83	2.85	2.86	2.94
	2.86	2.80	2.87	2.88

5.8 进行一个实验, 来研究操作温度和 3 种荧光屏玻璃类型对示波管的光输出的影响, 收集到的数据如

下：

玻璃类型	温度		
	100	125	150
1	580	1 090	1 392
	568	1 087	1 380
	570	1 085	1 386
2	550	1 070	1 328
	530	1 035	1 312
	579	1 000	1 299
3	546	1 045	867
	575	1 053	904
	599	1 066	889

(a) 在分析中用 $\alpha = 0.05$. 交互作用显著存在吗？玻璃类型或温度对响应有影响吗？你的结论是什么？

(b) 对于光输出，拟合一个合适的关于玻璃类型和温度的模型.

(c) 分析该实验的残差，并讨论你所考虑的模型的适合性.

5.9 考虑思考题 5.1 中的实验. 用合适的模型拟合响应数据，用这个模型提供关于过程操作条件的指导.

5.10 对于思考题 5.1 的数据，用 Tukey 检验法来确定压强因子的哪些水平有显著差异.

5.11 进行一项实验来确定究竟是燃烧温度还是炉膛位置会影响碳极的烘烤密度. 数据如下：

位置	温度 (℃)		
	800	825	850
1	570	1 063	565
	565	1 080	510
	583	1 043	590
2	528	988	526
	547	1 026	538
	521	1 004	532

假定不存在交互作用. 写出统计模型. 假定两个因子都是固定效应，进行方差分析. 能获得怎样的结论？论述模型的适合性.

5.12 设所有因子是固定的，对每单元一个观测值的二因子方差分析导出期望均方.

5.13 考虑下述二因子析因实验的数据. 分析这些数据并写出结论. 对非可加性进行检验. 用 $\alpha = 0.05$.

行因子	列因子			
	1	2	3	4
1	36	39	36	32
2	18	20	22	20
3	30	37	33	34

5.14 设想一种粘合剂的抗剪强度受应用压强和温度的影响. 设两个因子都是固定的，进行析因实验. 分析这些数据并写出结论. 对非可加性进行检验.

压强 ($\mathrm{lb/in^2}$)	温度 (℃)		
	250	260	270
120	9.60	11.28	9.00
130	9.69	10.10	9.57
140	8.43	11.01	9.03
150	9.98	10.44	9.80

*5.15 考虑三因子模型

$$y_{ijk}=\mu+\tau_i+\beta_j+\gamma_k+(\tau\beta)_{ij}+(\beta\gamma)_{jk}+\varepsilon_{ijk}\quad\begin{cases}i=1,2,\cdots,a\\j=1,2,\cdots,b\\k=1,2,\cdots,c\end{cases}$$

注意仅有一次重复. 假定所有因子都是固定的, 写出方差分析表, 包括期望均方. 为了检验假设, 你用什么作为“实验误差”?

5.16 研究对纸张强度有影响的因子. 它们是: 纸浆中硬木浓度的百分率、容器的压强以及煮浆时间. 选取 3 个硬木浓度水平、3 个压强水平、两个煮浆时间. 所有的因子都看作是固定效应. 进行有两次重复的析因实验, 得出下列数据:

硬木浓度百分率	煮浆时间 3 小时			煮浆时间 4 小时		
	压强			压强		
	400	500	650	400	500	600
2	196.6	197.7	199.8	198.4	199.6	200.6
	196.0	196.0	199.4	198.6	200.4	200.9
4	198.5	196.0	198.4	197.5	198.7	199.6
	197.2	196.9	197.6	198.1	198.0	199.0
8	197.5	195.6	197.4	197.6	197.0	198.5
	196.6	196.2	198.1	198.4	197.8	199.8

(a) 分析这些数据并写出结论. 用 $\alpha=0.05$.

(b) 作出恰当的残差图并论述模型的适合性.

(c) 你会在哪组条件下进行生产? 为什么?

*5.17 纺织厂的质量控制部门研究几个对用来做男士衬衫的棉花-合成纤维混纺布的染色有效应的因子. 选取 3 名操作工、3 种循环时间、两种温度, 以及在每组条件下着色布的 3 个小样品. 将成品与标准进行比较并打分. 结果如下. 分析这些数据并写出结论. 论述模型的适合性.

循环时间	温度					
	300 ℃			350 ℃		
	操作工			操作工		
	1	2	3	1	2	3
40	23	27	31	24	38	34
	24	28	32	23	36	36
	25	26	29	28	35	39
50	36	34	33	37	34	34
	35	38	34	39	38	36
	36	39	35	35	36	31
60	28	35	26	26	36	28
	24	35	27	29	37	26
	27	34	25	25	34	24

5.18 在思考题 5.1 中, 当任意两种压强产生的真实均差大于 0.5 时, 要以高概率拒绝零假设. 如果产率的标准差的一个合理的先验估计是 0.1, 应该进行多少次重复?

5.19 研究化学过程中的产率. 感兴趣的两个因子是温度和压强. 每个因子选择 3 个水平, 但是一天只能

进行 9 次试验. 实验者每天进行一次设计的完全重复. 得到下表的数据. 假定一天就是一个区组, 分析这些数据.

温度	第一天压强			第二天压强		
	250	260	270	250	260	270
低	86.3	84.0	85.8	86.1	85.2	87.3
中	88.5	87.3	89.0	89.4	89.9	90.3
高	89.1	90.2	91.3	91.7	93.2	93.7

5.20 考虑思考题 5.5 中的数据, 分析这些数据, 假定重复就是区组.

*5.21 考虑思考题 5.6 中的数据, 分析这些数据, 假定重复就是区组.

*5.22 在*Journal of Testing and Evaluation* (Vol. 16. no. 2, pp. 508~515) 上有篇论文, 研究了循环荷载和环境条件在 22 MPa 压力下对某种物质的疲劳裂纹增长的影响. 该实验的数据如下 (响应变量是裂纹增长率):

(a) 分析这个实验中的数据. 用 $\alpha = 0.05$.

(b) 分析残差.

(c) 用 $\ln(y)$ 作为响应变量, 重复 (a) 和 (b) 中的分析. 并论述这个结果.

频数	环境		
	空气	水	盐水
10	2.29	2.06	1.90
	2.47	2.05	1.93
	2.48	2.23	1.75
	2.12	2.03	2.06
1	2.65	3.20	3.10
	2.68	3.18	3.24
	2.06	3.96	3.98
	2.38	3.64	3.24
0.1	2.24	11.00	9.96
	2.71	11.00	10.01
	2.81	9.06	9.36
	2.08	11.30	10.40

5.23 在 *IEEE Transactions on Electron Devices* (Nov. 1986, pp. 1754) 有篇文章, 描述了对多晶硅掺杂的研究. 下表所示的实验是他们研究的一种变形. 响应变量是基极电流.

多晶硅掺杂 (ions)	热处理温度 (℃)		
	900	950	1 000
1×10^{20}	4.60	10.15	11.01
	4.40	10.20	10.58
2×10^{20}	3.20	9.38	10.81
	3.50	10.02	10.60

(a) 有无证据表明多晶硅掺杂或者热处理温度对基极电流有影响?

(b) 借助图形解释这个实验.

(c) 分析残差并论述模型的适合性.

(d) 这个实验支持模型 $y = \beta_0 + \beta_1 x_2 + \beta_2 x_2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$ 吗? 其中, x_1 代表掺杂水平, x_2 代表温度. 估计这个模型的参数并作出响应曲面.

5.24 研究两种不同品牌的电池在 3 种不同电器装置 (收音机、照相机、移动 DVD 播放机) 中的使用寿命 (单位: 小时). 进行一个完全随机的二因子析因实验, 得出以下数据:

电池品牌	电器装置		
	收音机	照相机	DVD 播放机
A	8.6	7.9	5.4
	8.2	8.4	5.7
B	9.4	8.5	5.8
	8.8	8.9	5.9

(a) 分析这些数据并写出结论 (用 α=0.05).

(b) 用残差图研究模型的适合性.

(c) 你会推荐哪种品牌的电池?

5.25 最近我新买了一些高尔夫球棒, 相信它们会有助于提高我的成绩. 下面是我在 3 个不同的高尔夫球场分别用新的和旧的球棒打出的 3 轮得分.

球棒	球　场		
	Ahwatukee	Karsten	Foothills
旧的	90	91	88
	87	93	86
	86	90	90
新的	88	90	86
	87	91	85
	85	88	88

(a) 取 α=0.05, 进行方差分析, 你能得出什么样的结论?

(b) 用残差图研究模型的适合性.

5.26 制造洗涤用品的厂商要研究新款去污剂的功效. 把新款去污剂和旧款去污剂用在有西红柿污点的棉布测试物上进行析因试验. 其他因子是测试物清洗次数 (1 次或 2 次) 和是否要借助清洁辅助物. 响应变量是在清洁后布料上残留污点的痕迹 (12 是最深, 0 是最浅). 数据如下表:

配方	清洗次数		清洗次数	
	1		2	
	辅助物		辅助物	
	是	否	是	否
新	6,5	6,5	3,2	4,1
旧	10,9	11,11	10,9	9,10

(a) 取 α=0.05, 进行方差分析, 你能得出什么样的结论?

(b) 用残差图研究模型的适合性.

第 6 章　2^k 析因设计

本章纲要

6.1　引言
6.2　2^2 设计
6.3　2^3 设计
6.4　一般的 2^k 设计
6.5　2^k 设计的单次重复
6.6　附加中心点的 2^k 设计
6.7　使用规范化设计变量的理由

第 6 章补充材料
S6.1　因子效应估计是最小二乘估计
S6.2　计算因子效应的 Yates 方法
S6.3　方差对照的一点注解
S6.4　预测响应的方差
S6.5　用残差识别分散效应
S6.6　中心点对析因点的重复的比较
S6.7　用 t 检验进行 "纯二次" 弯曲性的检验

6.1　引　　言

析因设计广泛应用于涉及多因子的实验, 所以有必要去研究这些因子对响应的联合效应. 第 5 章提出了析因设计分析的一般方法. 然而, 一般析因设计有几种特殊情况很重要, 因为它们广泛应用于研究工作, 并且它们也是其他一些有重要实践价值的设计的基础.

这些特殊情况中最重要的一种是, 有 k 个因子, 每个因子仅有两个水平. 这些水平可以是定量的, 例如是两个温度、两个压强或者两个时间的值; 也可以是定性的, 例如两台机器、两位操作员、一个因子的 "高" 水平和 "低" 水平或者一个因子的出现和不出现. 这类设计的一个完全的重复需要 $2\times2\times\cdots\times2=2^k$ 个观测值并称之为 **2^k 析因设计**.

本章重点关注这一类极为重要的设计. 本章假定: (1) 因子是固定的; (2) 设计是完全随机化的; (3) 满足通常的正态性假定.

在实验工作的早期阶段, 可能有很多因子需要研究时, 2^k 设计就特别有用. 它只需最少的实验次数就可以研究完全析因设计的 k 个因子. 因此, 这些设计可以广泛地用于**因子筛选实验**(factor screening experiment).

因为每一个因子仅有两个水平, 所以我们假定响应在所选的因子水平范围内是近似线性的. 在许多因子筛选实验中, 当我们开始研究过程或系统时, 这通常是一个合理的假定. 6.6 节将给出一个用于检验这一假定的简单方法, 并讨论一旦违背它将会发生什么情况.

6.2　2^2 设计

2^k 序列的第一个设计是仅有两个因子 (比如 A 与 B), 每一因子都有两个水平的设计. 这种设计叫做 2^2**析因设计**. 因子的水平可以任意叫做 "低" 和 "高". 例如, 研究一个化学过程中反应物浓度和催化剂量对于转化作用 (产率) 的效应. 实验的目标是确定对这 5 个因子中的任一个进行的调整是否提高了产率. 设反应物浓度是因子 A, 它的两个水平是 15%和 25%. 催化剂是因

子 B, 用 2 磅催化剂表示高水平, 仅用 1 磅, 表示低水平重复 3 次, 因此要进行 12 次试验. 试验的次序是随机的, 所以这是一个**完全随机化实验**, 得到的数据如下:

因子		处理组合	重复			总和
A	B		I	II	III	
−	−	A 低,B 低	28	25	27	80
+	−	A 高,B 低	36	32	32	100
−	+	A 低,B 高	18	19	23	60
+	+	A 高,B 高	31	30	29	90

此设计的 4 个处理组合如图 6.1 所示. 为方便起见, 因子的效应用大写拉丁字母表示. 这样, "A" 就表示因子 A 的效应, "B" 就表示因子 B 的效应, "AB" 就表示 AB 的交互作用. 在 2^2 设计中, A 与 B 的低水平与高水平分别在 A 轴与 B 轴上以 "−" 和 "+" 表示. 于是, A 轴上的 "−" 代表浓度的低水平 (15%), "+" 代表浓度的高水平 (25%); 而 B 轴上的 "−" 代表催化剂的低水平, "+" 表示催化剂的高水平.

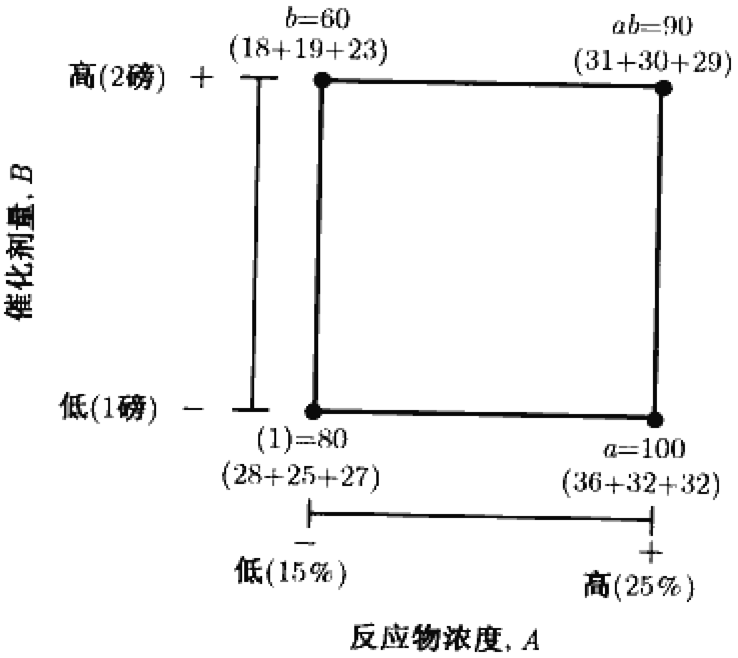


图 6.1 2^2 设计的处理组合

该设计中的 4 种处理组合通常用小写字母表示, 如图 6.1 所示. 从图中看出, 处理组合中任一因子的高水平可以用对应的小写字母表示, 而处理组合中任一因子的低水平用不写出对应的字母的方式来表示. 这样, a 代表 A 为高水平、B 为低水平的处理组合, b 代表 A 为低水平 B 为高水平的处理组合, ab 代表两个因子都是高水平的处理组合. 为方便起见, (1) 通常表示两个因子都是低水平的. 这种记法通用于 2^k 序列.

在二水平析因设计中, 一个因子的平均效应定义为那个因子的水平变化产生的响应变化在另一因子的水平上取平均值. 还有, 记号 (1), a, b, ab 现在代表那一处理组合取 n 次重复的总和, 如图 6.1 所说明的那样. 现在, A 在 B 的低水平上的效应为 $[a-(1)]/n$. A 在 B 的高水平上的效应为 $[ab-b]/n$. 取这两个量的平均值得 A 的主效应为:

$$A=\frac{1}{2n}\{[ab-b]+[a-(1)]\}=\frac{1}{2n}[ab+a-b-(1)] \qquad (6.1)$$

B 的平均主效应由 B 在 A 的低水平上的效应 (即, $[b-(1)]/n$) 和 B 在 A 的高水平上的效应 (即, $[ab-a]/n$) 求得:

$$B=\frac{1}{2n}\{[ab-a]+[b-(1)]\}=\frac{1}{2n}[ab+b-a-(1)] \tag{6.2}$$

交互作用效应 AB 定义为 A 在 B 的高水平上的效应与 A 在 B 的低水平上的效应之差的平均值. 于是得

$$AB=\frac{1}{2n}\{[ab-b]-[a-(1)]\}=\frac{1}{2n}[ab+(1)-a-b] \tag{6.3}$$

也可以定义 AB 为 B 在 A 的高水平上的效应和 B 在 A 的低水平上的效应之差的平均值. 这样定义也可以得出 (6.3) 式.

A, B, AB 的效应的计算公式也可以用另一方法推导出来. A 的效应可以由图 6.1 中的正方形右边的两个处理组合的平均响应 (称之为平均值 $\bar{y}_{A+}$, 因为它是在 A 的高水平处的处理组合的平均响应) 和左边的两个处理组合的平均响应 (即 $\bar{y}_{A-}$) 的差而求得. 也就是说,

$$A=\bar{y}_{A+}-\bar{y}_{A-}=\frac{ab+a}{2n}-\frac{b+(1)}{2n}=\frac{1}{2n}[ab+a-b-(1)]$$

这和 (6.1) 式的结果完全相同, B 的效应, 即 (6.2) 式, 可以由正方形上方两个处理组合的平均值 ($\bar{y}_{B+}$) 和正方形下方两个处理组合的平均值 ($\bar{y}_{B-}$) 的差而求得, 即

$$B=\bar{y}_{B+}-\bar{y}_{B-}=\frac{ab+b}{2n}-\frac{a+(1)}{2n}=\frac{1}{2n}[ab+b-a-(1)]$$

最后, 交互作用效应 AB 是正方形右至左对角线上的处理组合 [ab 与 (1)] 的平均值减去左至右对角线上的处理组合 (a 与 b) 的平均值, 即

$$AB=\frac{ab+(1)}{2n}-\frac{a+b}{2n}=\frac{1}{2n}[ab+(1)-a-b]$$

这与 (6.3) 式相同.

用图 6.1 的数据, 估计平均效应为

$$A=\frac{1}{2(3)}(90+100-60-80)=8.33$$
$$B=\frac{1}{2(3)}(90+60-100-80)=-5.00$$
$$AB=\frac{1}{2(3)}(90+80-100-60)=1.67$$

A 的效应 (反应物浓度) 是正的, 这表明它是递增的. A 从低水平 (15%) 增至高水平 (25%) 将增加产率. B 的效应 (催化剂) 是负的, 这表明在生产过程中增加催化剂的量会降低产率. 相对于两个主效应说来, 交互作用效应显得较小.

在很多涉及 2^k 设计的实验中, 我们将考察因子效应的**大小**(magnitude) 和**方向**(direction), 以便确定哪些变量可能是重要的. **方差分析**一般可用来证明这一点. 在建立和分析 2^k 设计的过程中, 有一些非常优秀的统计软件包很有用. 同时, 也有一些特殊的简便方法用来计算方差分析.

考虑 A, B, AB 的平方和. 由 (6.1) 式可得到一个用来估计 A 的对照, 即

$$\text{对照}_A = ab + a - b - (1) \tag{6.4}$$

通常称此对照为 A 的**总效应**. 由 (6.2) 式与 (6.3) 式可以看出，类似的对照亦用在估计 B 和估计 AB 中. 而且这 3 个对照是**正交的**. 任一对照的平方和可用 (3.29) 式来计算, 即对照的平方和等于对照的平方除以对照中的观测值的总个数乘对照系数的平方和. 因此有

$$SS_A = \frac{[ab + a - b - (1)]^2}{4n} \tag{6.5}$$

$$SS_B = \frac{[ab + b - a - (1)]^2}{4n} \tag{6.6}$$

$$SS_{AB} = \frac{[ab + (1) - a - b]^2}{4n} \tag{6.7}$$

它们分别是 A, B, AB 的平方和.

用图 6.1 的数据, 由 (6.5) 式、(6.6) 式和 (6.7) 式求得平方和为

$$\begin{aligned} SS_A &= \frac{(50)^2}{4(3)} = 208.33 \\ SS_B &= \frac{(-30)^2}{4(3)} = 75.00 \\ SS_{AB} &= \frac{(10)^2}{4(3)} = 8.33 \end{aligned} \tag{6.8}$$

用通常的方法求得总平方和, 即

$$SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{\cdots}^2}{4n} \tag{6.9}$$

一般地, SS_T 有 $4n-1$ 个自由度. 误差平方和有 $4(n-1)$ 个自由度, 用减法算得为

$$SS_E = SS_T - SS_A - SS_B - SS_{AB} \tag{6.10}$$

对于图 6.1 的实验, 用 (6.8) 式的 SS_A, SS_B, SS_{AB} 得

$$SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{3} y_{ijk}^2 - \frac{y_{\cdots}^2}{4(3)} = 9\ 398.00 - 9\ 075.00 = 323.00$$

$$SS_E = SS_T - SS_A - SS_B - SS_{AB} = 323.00 - 208.33 - 75.00 - 8.33 = 31.34$$

完整的方差分析概括在表 6.1 中. 根据 P 值, 我们得出结论: 两个主效应在统计上是显著的, 而因子间没有交互作用. 这一点肯定了我们原先根据因子效应的大小对数据所作的解释.

表 6.1　图 6.1 中的实验的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
A	208.33	1	208.33	53.15	0.000 1
B	75.00	1	75.00	19.13	0.002 4
AB	8.33	1	8.33	2.13	0.182 6
误差	31.34	8	3.92		
总和	323.00	11			

按顺序 (1), a, b, ab 写出处理组合常常很方便. 称这一顺序为**标准顺序**(或由 Frank Yates 提出的 Yates 顺序). 用标准顺序, 得出用来估计效应的对照系数是

效应	(1)	a	b	ab
A:	-1	$+1$	-1	$+1$
B:	-1	-1	$+1$	$+1$
AB:	$+1$	-1	-1	$+1$

注意, 估计交互作用效应的对照系数恰是两个主效应对应的系数的乘积. 对照系数不是 +1 就是− 1, 如表 6.2 列出的加和减符号表, 可用来确定每一种处理组合的固有符号. 表 6.2 各列的开头是主效应 (A 和 B)、AB 交互作用和 I(I 代表了整个实验的总和或平均值). 注意, 对应于 I 的那一列只有加号. 行的命名符号是处理组合. 为求出用来估计任一效应的对照, 只要用处理组合乘以表中相应于该效应的列的对应符号并加起来即可. 例如, 要估计 A, 其对照是 $-(1)+a-b+ab$, 与 (6.1) 式所得到的一致. 效应 A, B, AB 的对照是**正交**的. 所以, 2^2(乃至所有的 2^k 设计) 是**正交设计**.

表 6.2 计算 2^2 设计的效应的代数符号

处理组合	析因效应			
	I	A	B	AB
(1)	+	−	−	+
a	+	+	−	−
b	+	−	+	−
ab	+	+	+	+

1. 回归模型

在 2^k 析因设计中, 利用**回归模型**表达实验的结果是比较容易的. 因为 2^k 仅是一个析因设计, 我们既可以用效应模型, 也可以用均值模型, 但是回归模型方法更加自然和直接. 对于图 6.1 的化学过程实验, 回归模型是

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\varepsilon$$

其中 x_1 是代表反应物浓度的**规范变量**, x_2 是代表催化剂量的规范变量, β 是回归系数. **自然变量**, 即反应物浓度和催化剂量, 与规范变量之间的关系是

$$x_1=\frac{\text{浓度}-(\text{浓度}_{\text{低}}+\text{浓度}_{\text{高}})/2}{(\text{浓度}_{\text{高}}-\text{浓度}_{\text{低}})/2}$$

$$x_2=\frac{\text{催化剂}-(\text{催化剂}_{\text{低}}+\text{催化剂}_{\text{高}})/2}{(\text{催化剂}_{\text{高}}-\text{催化剂}_{\text{低}})/2}$$

当自然变量仅有两个水平时, 这种规范化对规范变量的水平产生熟悉的 ±1 记号. 用我们的例子说明这一点, 有

$$x_1=\frac{\text{浓度}-(15+25)/2}{(25-15)/2}=\frac{\text{浓度}-20}{5}$$

这样一来, 当浓度处于高水平 (浓度 =25%) 时, 则 $x_1 = +1$; 当浓度处于低水平 (浓度 =15%) 时, 则 $x_1 = -1$. 又,

$$x_2 = \frac{\text{催化剂} - (1+2)/2}{(2-1)/2} = \frac{\text{催化剂} - 1.5}{0.5}$$

当催化剂处于高水平 (催化剂 =2 磅) 时, $x_2 = +1$; 当催化剂处于低水平 (催化剂 =1 磅) 时, $x_2 = -1$.

拟合回归模型是

$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)x_1 + \left(\frac{-5.00}{2}\right)x_2$$

其中, 截距是所有 12 个观测值的总平均值, 回归系数 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 是相应因子效应估计量的一半. 回归系数是效应估计量的一半的理由是, 回归系数度量 x 上的一个单位的变化在 y 的均值上的效应, 而效应估计量是建立在两个单位变化 (从 -1 到 $+1$) 的基础上的. 估计回归系数的简单方法是进行**最小二乘**参数估计. 见本章的**补充材料**.

2. *残差与模型适合性*

该回归模型可用来产生 y 在本设计中的 4 个点处的预测与拟合值. 例如, 当反应物浓度处于低水平 ($x_1 = -1$) 以及催化剂处于低水平 ($x_2 = -1$) 时, 预测值是

$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)(-1) + \left(\frac{-5.00}{2}\right)(-1) = 25.835$$

在此处理组合处有 3 个观测值, 其残差是

$$e_1 = 28 - 25.835 = 2.165, \quad e_2 = 25 - 25.835 = -0.835, \quad e_3 = 27 - 25.835 = 1.165$$

同理, 可计算其余的预测值和残差. 对反应物浓度处于高水平而催化剂处于低水平时, 有

$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)(+1) + \left(\frac{-5.00}{2}\right)(-1) = 34.165$$

$$e_4 = 36 - 34.165 = 1.835, \quad e_5 = 32 - 34.165 = -2.165, \quad e_6 = 32 - 34.165 = -2.165$$

对反应物浓度的低水平和催化剂的高水平, 有

$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)(-1) + \left(\frac{-5.00}{2}\right)(+1) = 20.835$$

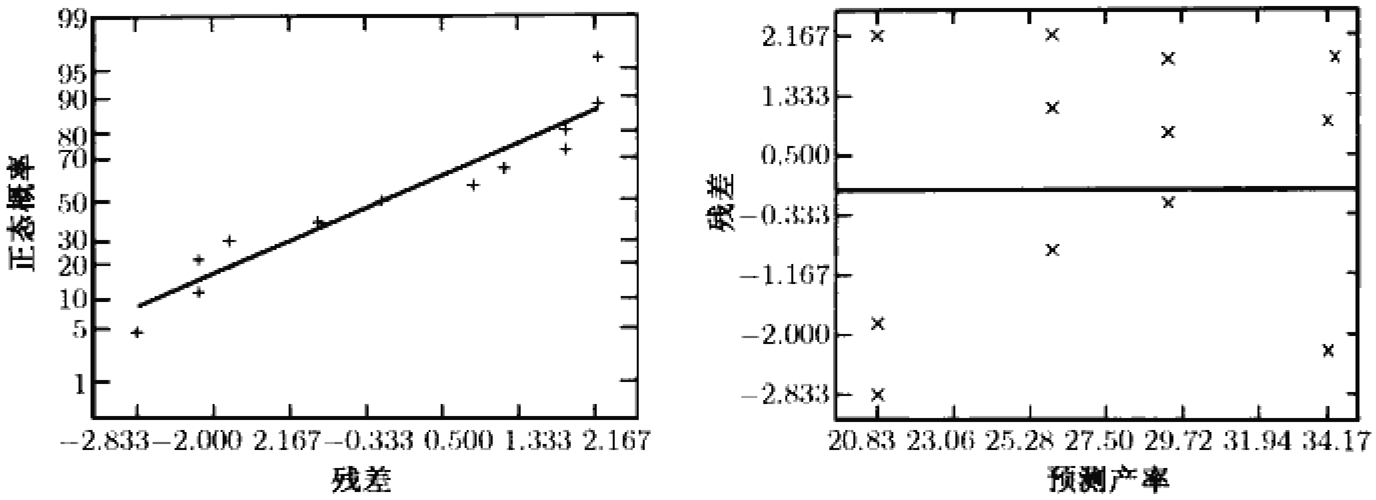
$$e_7 = 18 - 20.835 = -2.835, \quad e_8 = 19 - 20.835 = -1.835, \quad e_9 = 23 - 20.835 = 2.165$$

最后, 对两个因子的高水平, 有

$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)(+1) + \left(\frac{-5.00}{2}\right)(+1) = 29.165$$

$$e_{10} = 31 - 29.165 = 1.835, \quad e_{11} = 30 - 29.165 = 0.835, \quad e_{12} = 29 - 29.165 = -0.165$$

图 6.2 给出了这些残差的正态概率图以及残差与产率预测值的关系图. 这些图形是令人满意的, 所以, 没有理由怀疑我们的结论的有效性.



(a) 正态概率图　　(b) 残差与预测产率的关系图

图 6.2　化学过程实验的残差图

3. 响应曲面

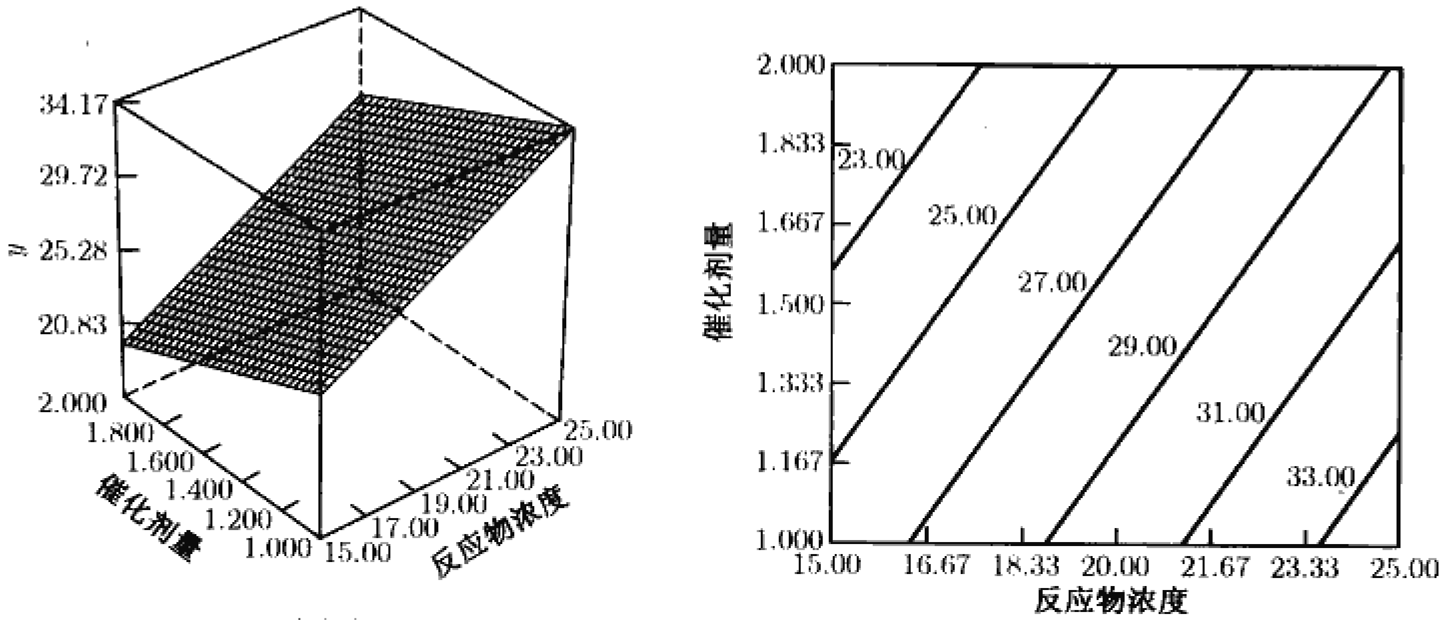
回归模型

$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)x_1 + \left(\frac{-5.00}{2}\right)x_2$$

可以用于生成响应曲面图. 如果要求用**自然因子水平**作图, 则可以简单地将早先得到的自然和规范变量之间的关系代入回归模型, 有

$$\begin{aligned}\hat{y} &= 27.5 + \left(\frac{8.33}{2}\right)\left(\frac{\text{浓度} - 20}{5}\right) + \left(\frac{-5.00}{2}\right)\left(\frac{\text{催化剂} - 1.5}{0.5}\right) \\ &= 18.33 + 0.833\ 3\text{浓度} - 5.00\text{催化剂}\end{aligned}$$

图 6.3a 给出了该模型的产率的三维响应曲面图, 图 6.3b 是等高线图. 因为模型是**一阶的**(也就是, 它只包含主效应), 拟合的响应曲面是一个平面. 通过检查等高线图, 可以看到产率可以随着反应物浓度的增加而提高, 随着催化剂量的减少而提高. 通常, 我们通过该拟合的曲面来确定一个过程的**潜在改进方向**. 一个被称为**最速上升法**的正规方法将在系统地研究响应曲面的第 11 章中介绍.

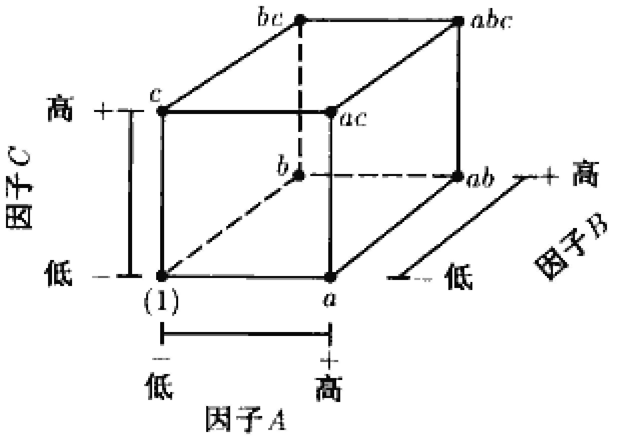


(a) 响应曲面　　(b) 等高线图

图 6.3　化学过程实验的产率响应曲面图与等高线图

6.3　2^3 设计

设有 3 个因子, A, B, C, 每一因子有两个水平. 此设计叫做 2^3**析因设计**. 现在, 如图 6.4a 所示, 在一个立方体上显示出 8 个处理组合. 用 “−” 和 “+” 的记号分别代表因子的低水平和高水平, 我们在图 6.5b 中列出 2^3 设计的 8 个实验. 有时我们称为**设计矩阵**. 推广 6.2 节所用的记号, 将处理组合依标准顺序写为 (1), a, b, ab, c, ac, bc, abc. 这些符号亦代表在所指定的处理组合上的 n 个观测值的总和.



(a) 几何观点

试验	因子		
	A	B	C
1	−	−	−
2	+	−	−
3	−	+	−
4	+	+	−
5	−	−	+
6	+	−	+
7	−	+	+
8	+	+	+

(b) 设计矩阵

图 6.4　2^3 析因设计

实际上有 3 种不同的记号广泛应用于 2^k 设计的实验中. 首先是 “+” 与 “−” 记号, 常称之为**几何记号**(geometric notation). 其次是用小写字母表示处理组合. 最后, 用记号 1 和 0 分别表示因子的高水平和低水平, 以代替 “+” 和 “−”. 对于 2^3 设计, 这些不同的记号说明如下:

试验	A	B	C	标签	A	B	C
1	−	−	−	(1)	0	0	0
2	+	−	−	a	1	0	0
3	−	+	−	b	0	1	0
4	+	+	−	ab	1	1	0
5	−	−	+	c	0	0	1
6	+	−	+	ac	1	0	1
7	−	+	+	bc	0	1	1
8	+	+	+	abc	1	1	1

2^3 设计的 8 个处理组合间有 7 个自由度. 3 个自由度和 A, B, C 的主效应有关. 4 个自由度与交互作用有关: AB, AC, BC 各有 1 个, ABC 有 1 个.

考虑估计主效应. 首先考虑估计主效应 A. 当 B 与 C 处于低水平时, A 的效应是 $[a-(1)]/n$. 同理, 当 B 处于高水平、C 处于低水平时, A 的效应是 $[ab-b]/n$. 当 C 处于高水平、B 处于低水平时, A 的效应是 $[ac-c]/n$. 最后, 当 B 和 C 都处于高水平时, A 的效应是 $[abc-bc]/n$. 这样一来, A 的平均效应正是这 4 个效应的平均值, 即

$$A=\frac{1}{4n}[a-(1)+ab-b+ac-c+abc-bc] \tag{6.11}$$

此式也可以用图 6.5a 的立方体右边一面的 4 个处理组合 (其中 A 处于高水平) 和左边一面的 4 个处理组合 (其中 A 处于低水平) 之间的对照推导出来. 也就是说, A 效应恰好是 A 处于高水平时 4 个试验的平均值 ($\bar{y}_{A+}$) 减去 A 处于低水平时 4 个试验的平均值 ($\bar{y}_{A-}$), 即

$$A = \bar{y}_{A+} - \bar{y}_{A-} = \frac{a + ab + ac + abc}{4n} - \frac{(1) + b + c + bc}{4n}.$$

此式可以重新排为

$$A = \frac{1}{4n}[a + ab + ac + abc - (1) - b - c - bc]$$

它与 (6.11) 式相同.

同理, B 的效应是立方体前边一面的 4 个处理组合的平均值和后边一面的 4 个处理组合的平均值之差, 即

$$B = \bar{y}_{B+} - \bar{y}_{B-} = \frac{1}{4n}[b + ab + bc + abc - (1) - a - c - ac] \tag{6.12}$$

C 的效应是立方体上边一面的 4 个处理组合的平均值和下边一面的 4 个处理组合的平均值之差, 即

$$C = \bar{y}_{C+} - \bar{y}_{C-} = \frac{1}{4n}[c + ac + bc + abc - (1) - a - b - ab] \tag{6.13}$$

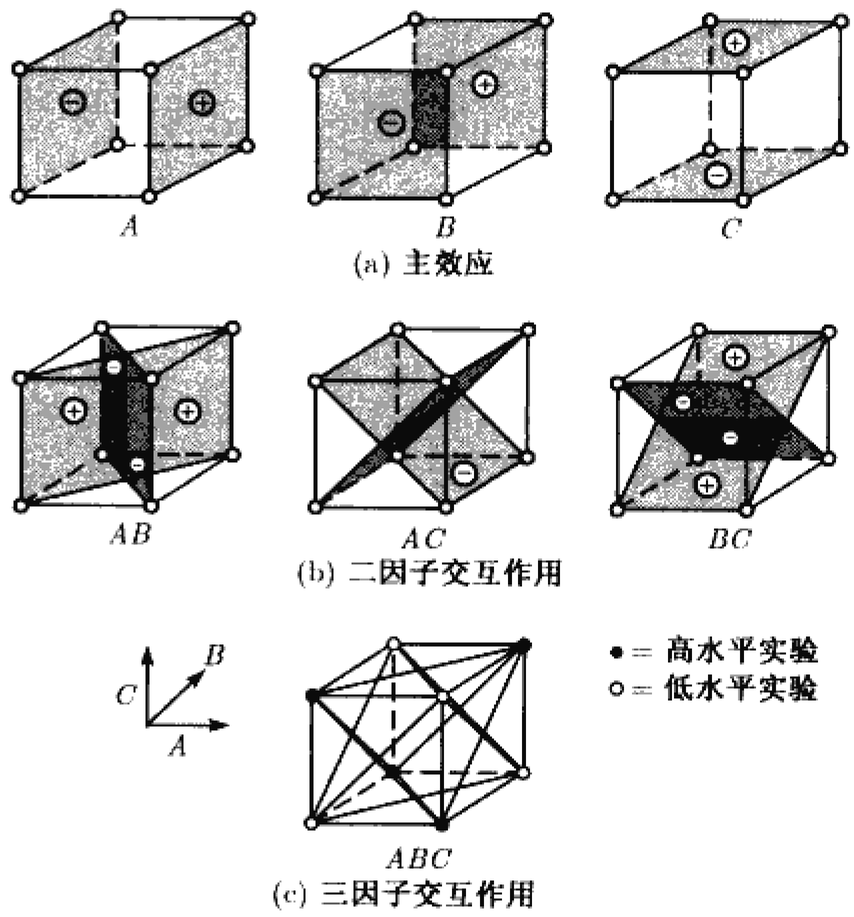


图 6.5 2^3 设计中对应于主效应和交互作用的对照的几何表示

二因子交互作用效应容易计算出来. AB 交互作用的一种度量是 A 在 B 的两个水平上的平均效应之差. 为方便起见, 称此差的一半为 AB 交互作用. 用符号表示为,

B	A 的平均效应
高 (+)	$\dfrac{[(abc-bc)+(ab-b)]}{2n}$
低 (−)	$\dfrac{\{(ac-c)+(a-(1))\}}{2n}$
差	$\dfrac{[abc-bc+ab-b-ac+c-a+(1)]}{2n}$

因为 AB 交互作用是此差的一半, 所以

$$AB=\frac{[abc-bc+ab-b-ac+c-a+(1)]}{4n} \tag{6.14}$$

(6.14) 式可写为：

$$AB=\frac{abc+ab+c+(1)}{4n}-\frac{bc+b+ac+a}{4n}$$

从这种形式容易看出, AB 交互作用就是图 6.5b 的立方体中在两个对角面上的试验的平均值之差. 同理并参考图 6.5b, 得 AC 和 BC 交互作用分别是

$$AC=\frac{1}{4n}[(1)-a+b-ab-c+ac-bc+abc] \tag{6.15}$$

$$BC=\frac{1}{4n}[(1)+a-b-ab-c-ac+bc+abc] \tag{6.16}$$

ABC 交互作用定义为 AB 在 C 的两个不同水平上的交互作用之差的平均值. 于是,

$$\begin{aligned}ABC&=\frac{1}{4n}\{[abc-bc]-[ac-c]-[ab-b]+[a-(1)]\}\\&=\frac{1}{4n}[abc-bc-ac+c-ab+b+a-(1)]\end{aligned} \tag{6.17}$$

和前面一样, 可以把 ABC 交互作用看作为两个平均值之差. 如果把两个平均值中的试验分离为两组, 那么, 它们就是图 6.5c 的立方体中的两个四面体的顶点.

在 (6.11) 式至 (6.17) 式中, 方括号中的量是处理组合的**对照**. 一张加减符号表可以从这些对照中导出来, 如表 6.3 所示. 主效应的符号是：高水平取加号, 低水平取减号. 一旦主效应的符号确定, 其余各列的符号可以逐行地用前面恰当的列的符号相乘而得. 例如, AB 列的符号就是 A 列的符号与 B 列的符号逐行的乘积. 任一效应的对照容易由此表求得.

表 6.3　计算 2^3 设计的效应的代数符号

处理组合	析因效应							
	I	A	B	AB	C	AC	BC	ABC
(1)	+	−	−	+	−	+	+	−
a	+	+	−	−	−	−	+	+
b	+	−	+	−	−	+	−	+
ab	+	+	+	+	−	−	−	−
c	+	−	−	+	+	−	−	+
ac	+	+	−	−	+	+	−	−
bc	+	−	+	−	+	−	+	−
abc	+	+	+	+	+	+	+	+

表 6.3 有几个有趣的性质：(1) 除列 I 之外, 每列加号的个数与减号的个数相等. (2) 任意两列符号乘积的和为零. (3) 列 I 与任一列相乘, 该列的符号不变. 也就是说, I 是一个单位元素(identity element). (4) 任意两列相乘, 得出表中的一列. 例如, $A \times B = AB$, 以及

$$AB \times B = AB^2 = A$$

我们看到, 乘积中的指数可以利用模为 2 的算术运算法求得. (也就是说, 指数仅是 0 或 1; 如果大于 1, 就将它减少 2 的某个倍数, 直至它为 0 或 1 为止.) 所有这些性质都包含在用来估计效应的对照的**正交性**中.

因为每一效应都有一个相应的单自由度对照, 所以效应的平方和是容易计算的. 在有 n 次重复的 2^3 设计中, 任一效应的平方和是

$$SS = \frac{(\text{对照})^2}{8n} \tag{6.18}$$

例 6.1 用 2^3 析因设计在单晶片等离子蚀刻工具中开发氮化物蚀刻工序. 设计因子包括电极间的间隙、气流 (C_2F_6 作为反应物气体) 和应用在阴极的 RF 功率 (见图 3.1 的晶片蚀刻工具示意图). 每个因子都有两个水平, 设计重复两次. 响应变量是硅氮化物的蚀刻率 (Å/m). 表 6.4 中显示的是蚀刻率数据. 图 6.6 几何地表示了该设计.

表 6.4 例 6.1 的晶片蚀刻实验

试验	规范因子			蚀刻率		总和	因子水平		
	A	B	C	重复 1	重复 2			低 (−1)	高 (+1)
1	−1	−1	−1	550	604	(1)=1 154	A(间隙, cm)	0.80	1.20
2	1	−1	−1	669	650	a = 1 319	B(C_2F_6 流, SCCM)	125	200
3	−1	1	−1	633	601	b = 1 234	C(功率, W)	275	325
4	1	1	−1	642	635	ab = 1 277			
5	−1	−1	1	1 037	1 052	c = 2 089			
6	1	−1	1	749	868	ac = 1 617			
7	−1	1	1	1 075	1 063	bc = 2 138			
8	1	1	1	729	860	abc = 1 589			

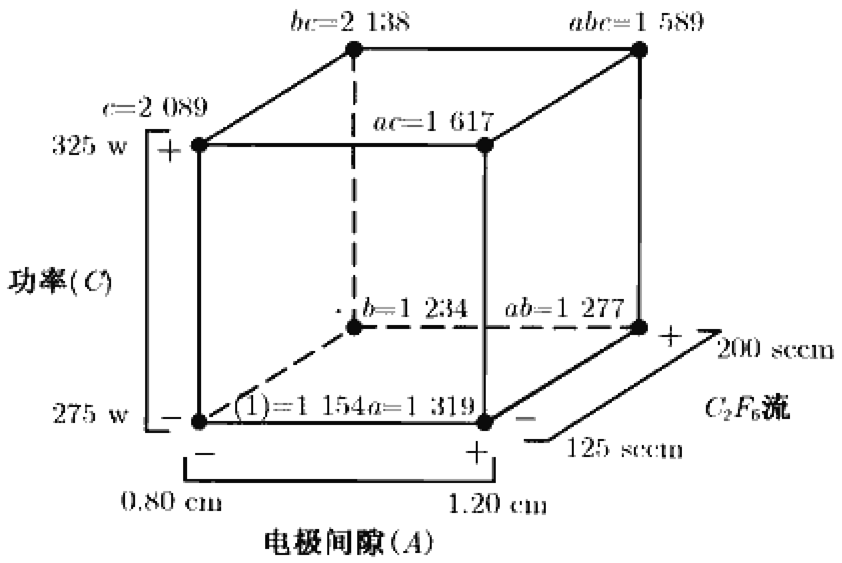


图 6.6 例 6.1 晶片蚀刻实验的 2^3 设计

用表 6.4 的处理组合的总和, 估计因子效应如下：

$$A = \frac{1}{4n}[a - (1) + ab - b + ac - c + abc - bc]$$
$$= \frac{1}{8}[1\ 319 - 1\ 154 + 1\ 277 - 1\ 234 + 1\ 617 - 2\ 089 + 1\ 589 - 2\ 138]$$

$$= \frac{1}{8}[-813] = -101.625$$

$$B = \frac{1}{4n}[b + ab + bc + abc - (1) - a - c - ac]$$

$$= \frac{1}{8}[1\ 234 + 1\ 277 + 2\ 138 + 1\ 589 - 1\ 154 - 1\ 319 - 2\ 089 - 1\ 617]$$

$$= \frac{1}{8}[59] = 7.375$$

$$C = \frac{1}{4n}[c + ac + bc + abc - (1) - a - b - ab]$$

$$= \frac{1}{8}[2\ 089 + 1\ 617 + 2\ 138 + 1\ 589 - 1\ 154 - 1\ 319 - 1\ 234 - 1\ 277]$$

$$= \frac{1}{8}[2\ 449] = 306.125$$

$$AB = \frac{1}{4n}[ab - a - b + (1) + abc - bc - ac + c]$$

$$= \frac{1}{8}[1\ 277 - 1\ 319 - 1\ 234 + 1\ 154 + 1\ 589 - 2\ 138 - 1\ 617 + 2\ 089]$$

$$= \frac{1}{8}[-199] = -24.875$$

$$AC = \frac{1}{4n}[(1) - a + b - ab - c + ac - c + abc]$$

$$= \frac{1}{8}[1\ 154 - 1\ 319 + 1\ 234 - 1\ 277 - 2\ 089 + 1\ 617 - 2\ 138 + 1\ 589]$$

$$= \frac{1}{8}[-1\ 229] = -153.625$$

$$BC = \frac{1}{4n}[(1) + a - b - ab - c - ac + bc + abc]$$

$$= \frac{1}{8}[1\ 154 + 1\ 319 - 1\ 234 - 1\ 277 - 2\ 089 - 1\ 617 + 2\ 138 + 1\ 589]$$

$$= \frac{1}{8}[-17] = -2.125$$

$$ABC = \frac{1}{4n}[abc - bc - ac + c - ab + b + a - (1)]$$

$$= \frac{1}{8}[1\ 589 - 2\ 138 - 1\ 617 + 2\ 089 - 1\ 277 + 1\ 234 + 1\ 319 - 1\ 154]$$

$$= \frac{1}{8}[45] = 5.625$$

最大的效应是功率 ($C = 306.125$)、间隙 ($A = -101.625$), 功率–间隙交互作用 ($AC = -153.625$).

由 (6.18) 式计算得平方和如下:

$$SS_A = \frac{(-813)^2}{16} = 41\ 310.562\ 5, \qquad SS_B = \frac{(59)^2}{16} = 217.562\ 5$$

$$SS_C = \frac{(2\ 449)^2}{16} = 374\ 850.062\ 5, \qquad SS_{AB} = \frac{(-199)^2}{16} = 2\ 475.062\ 5$$

$$SS_{AC} = \frac{(-1\ 229)^2}{16} = 94\ 402.562\ 5, \qquad SS_{BC} = \frac{(-17)^2}{16} = 18.062\ 5$$

$$SS_{ABC} = \frac{(45)^2}{16} = 126.562\ 5$$

总平方和是 SS_T=531 420.937 5, 用减法得 SS_E=18 020.50. 表 6.5 中汇总了效应估计及平方和. 标为"百分比贡献率"的列度量了模型的每一项对总的平方和的贡献百分比. 贡献百分比通常是粗略的, 但它是模型每一项相对重要性的有效指示. 注意, 主效应 C(功率) 事实上主导了该工序, 因为它占了总体变异性的 70%多, 而主效应 A(间隙) 和 AC 的交互作用各占大约 8%和 18%.

表 6.5 例 6.1 的效应估计描述

因子	效应估计	平方和	百分比贡献率
A	−101.625	41 310.562 5	7.773 6
B	7.375	217.562 5	0.040 9
C	306.125	374 850.062 5	70.537 3
AB	−24.875	2 475.062 5	0.465 7
AC	−153.625	94 402.562 5	17.764 2
BC	−2.125	18.062 5	0.003 4
ABC	5.625	126.562 5	0.023 8

表 6.6 中的方差分析可用于确定这些效应的大小. 它表明间隙和功率的主效应是高度显著的 (两个 P 值都很小). AC 的交互作用也是高度显著的, 因此, 间隙和功率之间有强烈的交互作用.

表 6.6 晶片蚀刻实验的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
间隙 (A)	41 310.562 5	1	41 310.562 5	18.34	0.002 7
气流 (B)	217.562 5	1	217.562 5	0.10	0.763 9
功率 (C)	374 850.062 5	1	374 850.062 5	166.41	0.000 1
AB	2 475.062 5	1	2 475.062 5	1.10	0.325 2
AC	94 402.562 5	1	94 402.562 5	41.91	0.000 2
BC	18.062 5	1	18.062 5	0.01	0.930 8
ABC	126.562 5	1	126.562 5	0.06	0.818 6
误差	18 020.500 0	8	2 252.562 5		
总和	531 420.937 5	15			

1. 回归模型和响应曲面

预测蚀刻率的回归模型是

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_3x_3+\hat{\beta}_{13}x_1x_3=776.062\,5+\left(\frac{-101.625}{2}\right)x_1+\left(\frac{306.125}{2}\right)x_3+\left(\frac{-153.625}{2}\right)x_1x_3$$

其中规范变量 x_1 与 x_3 分别表示 A 与 C. x_1x_3 一项表示 AC 的交互作用. 残差由观测得到的蚀刻率和预测蚀刻率之间的差来求得. 我们把这些残差的分析作为练习留给读者.

图 6.7 给出了从回归模型得到的蚀刻率的响应曲面和等高线图. 因为模型含有交互作用, 蚀刻率的等高线是曲线 (响应曲面是"扭曲的"平面). 理想的工序要求蚀刻率接近 900Å /m. 等高线图表明间隙和功率的几个组合都满足这个目标. 但是, 精确地控制这些变量是有必要的.

2. 计算机处理

有许多统计软件包可以用于建立和分析二水平析因设计. 其中的一种计算机程序 (Design-Expert) 的输出列在表 6.7 中. 表的上半部列出的是**全模型**的方差分析表. 这里的格式与表 6.6 中的方差分析的结果有所不同. 方差分析的第一行是对全模型 (所有主效应和交互作用) 的总概

括, 模型的平方和为

$$SS_{\text{模型}} = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC} = 5.134 \times 10^5$$

统计量

$$F_0 = \frac{MS_{\text{模型}}}{MS_E} = \frac{73\ 342.92}{2\ 252.56} = 32.56$$

用于检验假设

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_{12} = \beta_{13} = \beta_{23} = \beta_{123} = 0$$
$$H_1: \text{至少有一个 } \beta \neq 0$$

因为 F_0 大, 我们得出结论: 至少一个变量有非零效应. 用 F 统计量检验每一个析因设计效应的显著性. 其结果与表 6.6 一致.

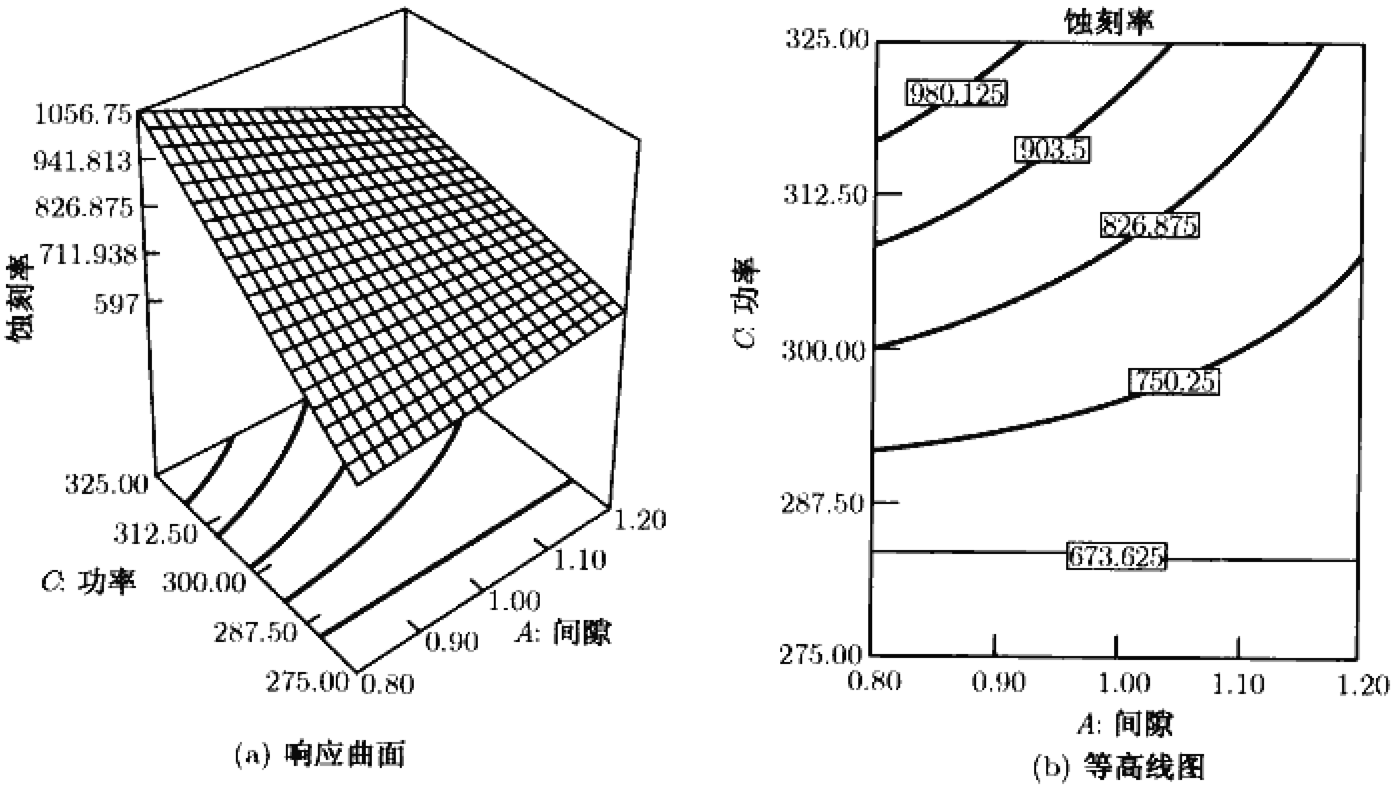


图 6.7 例 6.1 的蚀刻率的响应曲面和等高线图

表 6.7 中全模型方差分析的下面, 列出了几个 R^2 统计量. 普通的 R^2 为

$$R^2 = \frac{SS_{\text{模型}}}{SS_{\text{总和}}} = \frac{5.134 \times 10^5}{5.314 \times 10^5} = 0.966\ 1$$

它度量了由模型解释的总变异性的比例. 该统计量的一个问题是它随着模型因子个数的增加而增加, 即使这些因子是不显著的. *已调整的*R^2 统计量, 定义为

$$R^2_{\text{调整}} = 1 - \frac{SS_E/df_E}{SS_{\text{总和}}/df_{\text{总和}}} = 1 - \frac{18\ 020.50/8}{5.314 \times 10^5/15} = 0.936\ 4$$

它是根据模型的 "大小"(即因子个数) 调整的统计量. 如果模型中增加不显著项, 已调整 R^2 的实际上会随之降低. PRESS 统计量是衡量模型预测新数据精确程度的指标. (PRESS 是 Prediction Error Sum of Square 的缩写. 通过用除第 i 个以外, 其他观测值都在内的模型来预测第 i 个数

表 6.7 例 6.1 的 Design-Expert 输出

Response:Etch rate

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob> F
Model	5.134E+0.05	7	73342.92	32.56	< 0.0001
A	41310.56	1	41310.56	18.34	0.0027
B	217.56	1	217.56	0.097	0.7639
C	3.749E+005	1	3.749E+005	166.41	< 0.0001
AB	2475.06	1	2475.06	1.10	0.3252
AC	94402.56	1	94402.56	41.91	0.0002
BC	18.06	1	18.06	8.019E-003	0.9308
ABC	126.56	1	126.56	0.056	0.8186
Pure Error	18020.50	8	2252.56		
Cor Total	5.314E+005	15			
Std.Dev	47.46			R-Squared	0.9661
Mean	776.06			Adj R-Squared	0.9364
C.V.	6.12			Pred R-Squared	0.8644
PRESS	72082.00			Adeq Precisioin	14.660

Factor	Coefficient Estimated	DF	Standard Error	95%CI Low	95%CI High	VIF
Intercept	776.06	1	11.87	748.70	803.42	
A-Gap	−50.81	1	11.87	−78.17	−23.45	1.00
B-Gas flow	3.69	1	11.87	−23.67	31.05	1.00
C-Power	153.06	1	11.87	125.70	180.42	1.00
AB	−12.44	1	11.87	−39.80	14.92	1.00
AC	−76.81	1	11.87	−104.17	−49.45	1.00
BC	−1.06	1	11.87	−28.42	26.30	1.00
ABC	2.81	1	11.87	−24.55	30.17	1.00

Final Equation in Terms of Coded Factors:

Etchrate	=
+776.06	
−50.81	*A
+3.69	*B
+153.06	*C
−12.44	*A*B
−76.81	*A*C
+1.06	*B*C
+2.81	*A*B*C

Final Equation in Terms of Actual Factors:

Etch rate	=
−6487.33333	
+5355.41667	*Gap
+6.59667	*Gas flow
+24.10667	*Power
−6.15833	*Gap*Gas flow
−17.80000	*Gap*Power
−0.016133	*Gas flow*Power
+0.015000	*Gap*Gas flow*Power

(续)

Response:Etch rate

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob> F
Model	5.106E+005	3	1.702E+005	97.91	<0.0001
A	41310.56	1	41310.56	23.77	0.0004
C	3.749E+005	1	3.749E+005	215.66	<0.0001
AC	94402.56	1	94402.56	54.31	<0.0001
Residual	20857.75	12	1738.15		
Lack of Fit	2837.25	4	709.31	0.31	0.8604
Pure Error	18020.50	8	2252.56		
Cor Total	5.314E+005	15			

Std.Dev.	41.69	R-Squared	0.9608
Mean	776.06	Adj R-Squared	0.9509
C.V.	5.37	Pred R-Squared	0.9302
PRESS	37080.44	Adeq Precision	22.055

Facto	Coefficient Estimate	DF	Standard Error	95%CI Low	95%CI High	VIF
Intercept	776.06	1	10.42	753.35	798.77	
A-Gap	−50.81	1	10.42	−73.52	28.10	1.00
C-Power	153.06	1	10.42	130.35	175.77	1.00
AC	−76.81	1	10.42	−99.52	−54.10	1.00

Final Equation in Terms of Coded Factors:

Etch rate =
+776.06
−50.81 *A
+153.06 *C
−76.81 *A*C

Final Equation in Terms of Actual Factors:

Etch rate =
−5415.37500
+4354.68750 *Gap
+21.48500 *Power
−15.36250 *Gap*Power

Diagnostics Case Statistics

Standard Order	Actual Value	Predicted Value	Residual	Leverage	Student Residual	Cook's Distance	Outilier t	Run Order
1	550.00	597.00	−47.00	0.250	−1.302	0.141	−1.345	9
2	604.00	597.00	7.00	0.250	0.194	0.003	0.186	6
3	669.00	649.00	20.00	0.250	0.554	0.026	0.537	14
4	650.00	649.00	1.00	0.250	0.028	0.000	0.027	1
5	633.00	597.00	36.00	0.250	0.997	0.083	0.997	3
6	601.00	597.00	4.00	0.250	0.111	0.001	0.106	12
7	642.00	649.00	−7.00	0.250	−0.194	0.003	−0.186	13
8	635.00	649.00	−14.00	0.250	−0.388	0.013	−0.374	8
9	1037.00	1056.75	−19.75	0.250	−0.547	0.025	−0.530	5

(续)

Diagnostics Case Statistics								
Standard Order	Actual Value	Predicted Value	Residual	Leverage	Student Residual	Cook's Distance	Outilier t	Run Order
10	1052.00	1056.75	−4.75	0.250	−0.132	0.001	−0.126	16
11	749.00	801.50	−52.50	0.250	−1.454	0.176	−1.534	2
12	868.00	801.50	66.50	0.250	1.842	0.283	2.082	15
13	1075.00	1056.75	18.25	0.250	0.505	0.021	0.489	4
14	1063.00	1056.75	6.25	0.250	0.173	0.002	0.166	7
15	729.00	801.50	−72.50	0.250	−2.008	0.336	−2.359	10
16	860.00	801.50	58.50	0.250	1.620	0.219	1.755	11

据点, 这样所得到的预测误差平方的和可用来计算它.) PRESS 值小的模型表示是好的预测模型. "预测 R^2" 统计量由

$$R^2_{预测} = 1 - \frac{\text{PRESS}}{SS_{总和}} = 1 - \frac{72\ 082.00}{5\ 314 \times 10^5} = 0.864\ 4$$

计算. 这表明全模型大概可以解释新数据 86%的变异性.

输出的下一部分列出了每个模型项的回归系数以及每个系数的**标准误**, 其定义为

$$se(\hat{\beta}) = \sqrt{V(\hat{\beta})} = \sqrt{\frac{MS_E}{2^k n}} = \sqrt{\frac{2\ 252.56}{2(8)}} = 11.87$$

每个回归系数的 95%置信区间由

$$\hat{\beta} - t_{0.025,N-p}se(\hat{\beta}) \leqslant \beta \leqslant \hat{\beta} + t_{0.025,N-p}se(\hat{\beta})$$

计算得出, 其中 t 的自由度就是误差自由度的数目, 也就是, N 是实验中的试验次数 (16), p 是模型的参数个数 (8). 同时还列出了基于规范变量和自然变量的全模型.

表 6.7 中的最后一部分给出了去掉不显著项后的输出. 该**简化模型**仅包含主效应因子 A, C 以及 AC 的交互作用. 误差或残差平方和可分解为源于立方体的 8 个角点的重复的**纯误差**分量, 以及模型中去除因子 (B, AB, BC 及 ABC) 的平方和的**拟合不足**分量. 表中还列出了分别用规范变量和自然变量表示的回归模型. 模型解释的蚀刻率的总变异性比例为

$$R^2 = \frac{SS_{模型}}{SS_{总和}} = \frac{5.106 \times 10^5}{5.314 \times 10^5} = 0.960\ 8$$

它比全模型的 R^2 小. 但是简化模型的已调整的 R^2 比全模型的已调整的 R^2 稍大, 而简化模型的 PRESS 相当小, 使得简化模型的 R^2_{rep} 较大. 显然, 从全模型中去掉不显著项得到了最终模型, 它对预测新数据更有效. 注意到, 简化模型的回归系数的置信区间比相应的全模型的置信区间要短.

输出的最后部分列出了简化模型的残差. Design-Expert 也给出了前面讨论过的所有残差图.

1. 判断效应显著性的其他方法

方差分析是用来确定哪些因子效应不为零的一种正规方法. 还有几种其他的有用方法. 下面将介绍如何计算**效应的标准误**, 并利用这些标准误去构造效应的**置信区间**. 另一种方法将在 6.5 节中说明, 用**正态概率图**来评定效应的重要性.

易求一效应的标准误. 如果我们假定设计的 2^k 次试验的每一次都有 n 次重复, 而 $y_{i1}, y_{i2}, \cdots, y_{in}$ 是第 i 次试验的观测值, 则

$$S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_{ij}-\bar{y}_i)^2 \quad i=1,2,\cdots,2^k$$

是第 i 次试验的一个方差估计. 可以合并这 2^k 个方差以得到总方差估计:

$$S^2 = \frac{1}{2^k(n-1)}\sum_{i=1}^{2^k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_i)^2 \tag{6.19}$$

这也是方差分析中的用误差均方给出的方差估计. 每一效应的方差估计是

$$V(\text{效应}) = V\left(\frac{\text{对照}}{2^{k-1}n}\right) = \frac{1}{(2^{k-1}n)^2}V(\text{对照})$$

每一对照是 2^k 个处理总和的线性组合, 而每一总和是由 n 个观测值构成的. 因此,

$$V(\text{对照}) = 2^k n\sigma^2$$

且一效应的方差是

$$V(\text{效应}) = \frac{1}{(2^{k-1}n)^2}2^k n\sigma^2 = \frac{1}{2^{k-2}n}\sigma^2$$

将 σ^2 代以它的估计量 S^2 并取最后表达式的平方根就可求得标准误的估计:

$$se(\text{效应}) = \frac{2S}{\sqrt{2^k n}} \tag{6.20}$$

注意到, 2^k 设计中, 效应的标准误是回归模型中回归系数估计量的标准误的 2 倍. (见例 6.1 的 Design-Expert 计算机输出).

用效应 $\pm t_{\alpha/2,N-p}se(\text{效应})$ 计算效应的 $100(1-\alpha)\%$置信区间, 其中 t 的自由度是误差或残差的自由度 ($N-p=$ 试验总数 $-$ 模型参数的个数).

为说明这一方法, 考虑例 6.1 的晶片蚀刻实验. 全模型的均方误是 MS_E=2 252.56, 每一效应的标准误是 (用 $S^2 = MS_E$)

$$se(\text{效应}) = \frac{2S}{\sqrt{2^k n}} = \frac{2\sqrt{2\ 252.56}}{\sqrt{2(2^3)}} = 23.73$$

这时, $t_{0.025,8} = 2.31$, $t_{0.025,8}se(\text{效应}) = 2.31(23.73) = 54.82$, 于是, 各因子效应的近似 95%置信区间是

$$\begin{aligned} A:&\quad -101.625 \pm 54.82\\ B:&\quad 7.375 \pm 54.82\\ C:&\quad 306.125 \pm 54.82\\ AB:&\quad -24.875 \pm 54.82 \end{aligned}$$

$$AC: \quad -153.625 \pm 54.82$$
$$BC: \quad -2.125 \pm 54.82$$
$$ABC: \quad 5.625 \pm 54.82$$

分析表明 A, C, AC 都是重要的因子, 因为只有它们的近似 95%置信区间不包含零.

2. 分散效应

晶片蚀刻工具过程中的工程师也感兴趣于**分散效应**(dispersion effect), 也就是说, 进行试验时, 任一因子确实影响不同试验间蚀刻率的变异性吗? 回答这个问题的一种方法是, 对于 2^3 设计的 8 个试验的每一个, 观测其蚀刻率的*极差*(range). 这些极差画在图 6.8 的立方体上. 注意到当间隙和功率都处在高水平时, 蚀刻率的极差要大得多, 这表明这两个因子水平的这一组合可能会导致蚀刻率的变异性比其他搭配更大. 幸运的是, 通过设置间隙和功率, 可以达到蚀刻率要求的 900Å /m 左右, 从而避免这种情况.

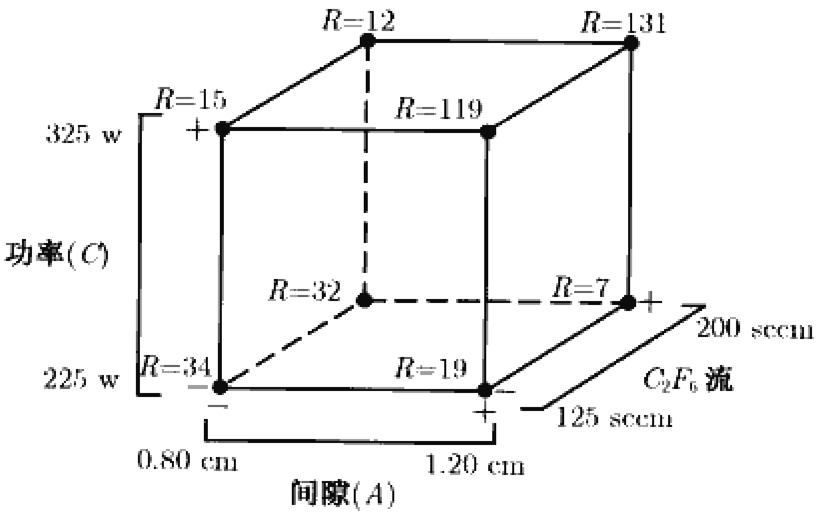


图 6.8 例 6.1 的蚀刻率的极差

6.4 一般的 2^k 设计

前面提出的分析方法可推广到 2^k 析因设计 (即有 k 个因子、每因子有两个水平) 的情形中去. 2^k 设计的统计模型包含 k 个主效应, $\binom{k}{2}$ 个二因子交互作用, $\binom{k}{3}$ 个三因子交互作用, ……, 以及一个 k 因子交互作用. 也就是说, 对 2^k 设计, 全模型含有 $2^k - 1$ 个效应. 前面对于处理组合而引入的记号也适用于此处. 例如, 在 2^5 设计中, abd 表示因子 A, B, D 处于高水平而因子 C 和 E 处于低水平的处理组合. 处理组合可以按**标准顺序**写出, 方法是, 每引入一个新的因子, 就依次和前面已引入的因子进行组合. 例如, 2^4 设计的标准顺序是 (1), $a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, abcd$.

2^k 设计的统计分析的一般方法概括在表 6.8 中. 正如前面指出的一样, 分析过程通常可以使用一个计算机软件包.

表 6.8 中的步骤相信大家都是熟悉的. 第一步是估计因子效应并检查它们的符号和大小. 这给出了实验者关于哪些因子和交互作用可能是重要的、在什么方向可以调整这些因子以改进响应的最初信息. 在形成实验的初始模型时, 通常选择全模型, 即所有主效应和交互作用, 只要至少一个设计点重复 (6.5 节会讨论这一步的修改). 第三步, 利用方差分析进行关于主效应和交互

作用的显著性检验. 表 6.9 给出了 n 次重复 2^k 析因设计的方差分析的一般形式. 第四步, 改进模型, 通常包括从全模型中移走不显著项. 第五步, 用通常的残差分析来检验模型适合性和假定. 如果模型不适合或严重偏离假定那么在残差分析之后, 我们有时会重新改进模型. 最后一步通常包括图形分析: 或主效应图或交互作用图或响应曲面和等高线图.

表 6.8　2^k 设计的分析步骤

(1) 估计因子效应	(4) 改进模型
(2) 形成初始模型	(5) 分析残差
(a) 如果设计是重复的, 则拟合全模型	(6) 解释结果
(b) 如果没有重复, 则用效应的正态概率图形成模型	
(3) 进行统计检验	

表 6.9　2^k 设计的方差分析

方差来源	平方和	自由度
k 主效应		
A	SS_A	1
B	SS_B	1
$\vdots$	$\vdots$	$\vdots$
K	SS_K	1
$\binom{k}{2}$ 二因子交互作用		
AB	SS_{AB}	1
AC	SS_{AC}	1
$\vdots$	$\vdots$	$\vdots$
JK	SS_{JK}	1
$\binom{k}{3}$ 三因子交互作用		
ABC	SS_{ABC}	1
ABD	SS_{ABD}	1
$\vdots$	$\vdots$	$\vdots$
IJK	SS_{IJK}	1
$\vdots$		
$\binom{k}{k}$ k 因子交互作用		
$ABC\cdots K$	$SS_{ABC\cdots K}$	1
误差	SS_E	$2^k(n-1)$
总和	SS_T	2^kn-1

尽管前面叙述的计算大多数可以利用计算机进行, 但有时手工计算效应估计或效应的平方和也是必要的. 要估计一个效应或计算一个效应的平方和, 必须先确定该效应的对照. 可以使用如表 6.2 或表 6.3 那样的加号和减号表做到这一点. 不过, 对大的 k 值, 那将很麻烦, 可以采用其他方法. 一般说来, 确定效应 $AB\cdots K$ 的对照可以用展开下式右边的方法:

$$(\text{对照})_{AB\cdots K}=(a\pm 1)(b\pm 1)\cdots(k\pm 1) \tag{6.21}$$

在展开 (6.21) 式时, 按初等代数方法计算, 而在最后的表示式中, 用 (1) 代替 "1". 小括号中的符号的选取法是: 当因子包含在效应中时, 取负号; 不在时, 取正号.

为说明 (6.21) 式的用法, 考虑 2^3 析因设计, AB 的对照是

$$(\text{对照})_{AB} = (a-1)(b-1)(c+1) = abc + ab + c + (1) - ac - bc - a - b$$

进一步, 在 2^5 设计中, $ABCD$ 的对照是

$$\begin{aligned}(\text{对照})_{ABCD} &= (a-1)(b-1)(c-1)(d-1)(e+1)\\ &= abcde + cde + bde + ade + bce\\ &\quad + ace + abe + e + abcd + cd + bd\\ &\quad + ad + bc + ac + ab + (1) - a - b - c\\ &\quad - abc - d - abd - acd - bcd - ae\\ &\quad - be - ce - abce - de - abde - acde - bcde\end{aligned}$$

一旦算出效应的对照, 就可根据下式估计效应并计算其平方和:

$$AB\cdots K = \frac{2}{2^k n}(\text{对照}_{AB\cdots K}) \tag{6.22}$$

$$SS_{AB\cdots K} = \frac{1}{2^k n}(\text{对照}_{AB\cdots K})^2 \tag{6.23}$$

其中 n 表示重复的次数. 也有基于 Frank Yates 的表格算法, 有时它对效应估计及平方和的手工计算是有用的. 详见本章的**补充材料**.

6.5 2^k 设计的单次重复

即使是对于中等大小的因子个数, 2^k 析因设计的处理组合的总数也是很大的. 例如, 2^5 设计有 32 个处理组合, 2^6 设计有 64 个处理组合等. 因为资源通常受限, 实验者能做的重复次数也就受到限制. 通常可用的资源只允许对设计做**一次重复**, 除非实验者愿意消去某些原来的因子.

每个试验组合仅进行一次实验的风险是, 我们可能拟合了含噪声的模型. 即, 若响应 y 变化程度大, 这可能会误导从实验中得到的结论. 如图 6.9a 所示. 该图中, 直线代表真实的因子效应. 但是, 由于出现在响应变量 (用阴影带表示) 中的随机变异性, 实验者得到了用黑圆点表示的两个测量到的响应. 于是, 因子效应的估计接近于零. 结果, 实验者得到了关于该因子的错误结论. 如果响应的变化程度较小, 错误结论的可能性也会较小. 另一种确保得到可靠的效应估计的方法是增加低 (−) 水平和高 (+) 水平因子间的距离, 如图 6.9b 所示. 该图中, 低因子水平和高因子水平间的增加了的距离得到了真实因子效应的合理估计.

一次重复策略通常用于有相对多的需考虑因子的筛选实验. 因为, 我们在不能完全确信实验误差小的情形下, 采取尽量拉开因子水平是一个好的做法. 重读第 1 章中选择因子水平的指南, 也许会有帮助.

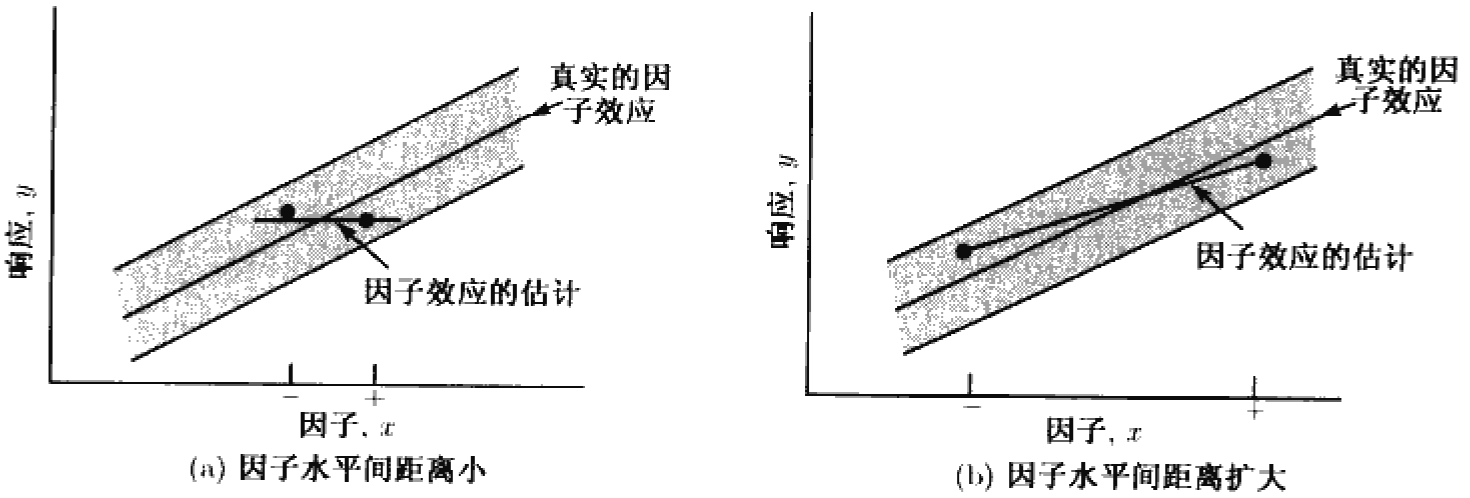


图 6.9 无重复设计的因子水平选择的影响

2^k 设计的单次重复有时称为**无重复的析因设计**. 因为仅有一次重复, 所以没有误差的内在估计 (即 "纯误差"). 无重复析因设计的一种分析法是, 假定某些高阶的交互作用可被忽略, 并将它们的均方组合起来用于估计误差. 这要借助于**效应稀疏原理**(sparsity of effect principle), 也就是, 很多系统的一些主效应和低阶的交互作用处于支配地位, 而很多高阶交互作用可被忽略.

当分析无重复析因设计的数据时, 有时出现真正的高阶交互作用. 此时, 将高阶交互作用集中起来作为误差均方使用是不恰当的. Daniel(1959) 提出的分析方法是克服这一困难的简便途径. Daniel 建议, 检查效应估计量的**正态概率图**. 可被忽略的效应是正态分布的, 其均值为零、方差为 σ^2, 因此会大致落在图上的一条直线附近; 而显著效应有非零均值因此不会落在这一直线上. 因此, 基于正态概率图, 初始模型必须指定包含那些显著不为零的效应. 显然可以将忽略的效应组合成误差的估计.

例 6.2 2^4 设计的单次重复

一种化学产品是在压力容器中生产的. 在一个小规模的试验性工厂中进行一项析因实验, 用以研究可能会影响这种产品渗透率的因子. 4 个因子是温度 (A)、压强 (B)、甲醛浓度 (C) 和搅拌速度 (D). 每一因子取两个水平, 2^4 试验的单次重复所得的数据如表 6.10 和图 6.10 所示. 依随机次序进行 16 次试验. 工程师感兴趣的是使渗透率达到最大. 在当前的生产条件下, 产品的渗透率约为 75 gal/h(加仑/小时). 当前生产用的甲醛浓度, 即因子 C, 为高水平. 工程师希望尽可能减少甲醛浓度, 但一直未能做到这一点, 因为这总是造成渗透率太低.

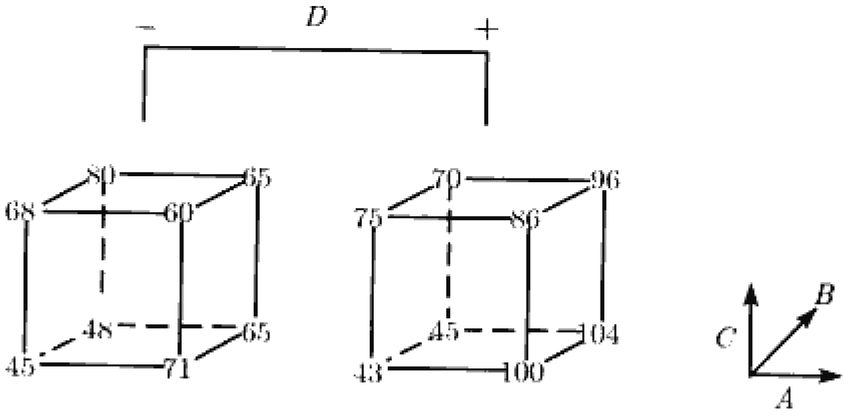


图 6.10 例 6.2 的试验工厂渗透率实验的数据

开始分析这些数据时, 先将效应估计量画在正态概率纸上. 2^4 设计的对照常数的加号和减号表如表 6.11 所示. 由这些对照, 可以估计 15 个因子效应以及平方和, 如表 6.12 所示.

这些效应的正态概率图如图 6.11 所示. 沿直线上的所有效应可被忽略, 而大的效应则远离此直线. 此

分析显露出的重要效应是主效应 A, C, D 以及 AC 交互作用和 AD 交互作用.

表 6.10 小规模试验工厂的渗透率实验

试验编号	因子				试验标号	渗透率 (gal/h)
	A	B	C	D		
1	−	−	−	−	(1)	45
2	+	−	−	−	a	71
3	−	+	−	−	b	48
4	+	+	−	−	ab	65
5	−	−	+	−	c	68
6	+	−	+	−	ac	60
7	−	+	+	−	bc	80
8	+	+	+	−	abc	65
9	−	−	−	+	d	43
10	+	−	−	+	ad	100
11	−	+	−	+	bd	45
12	+	+	−	+	abd	104
13	−	−	+	+	cd	75
14	+	−	+	+	acd	86
15	−	+	+	+	bcd	70
16	+	+	+	+	$abcd$	96

表 6.11 2^4 设计的对照常数

	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	$ABCD$
(1)	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+
a	+	−	−	−	−	+	+	−	−	+	+	+	+	−	−
b	−	+	−	−	+	−	+	−	+	−	+	+	−	+	−
ab	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+
c	−	−	+	+	−	−	+	−	+	+	−	−	+	+	−
ac	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+
bc	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+
abc	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−
d	−	−	+	−	+	+	−	+	−	−	+	−	+	+	−
ad	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+
bd	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+
abd	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−
cd	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+
acd	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−
bcd	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−
$abcd$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

图 6.12a 画出了主效应 A, C, D 的图形. 所有这 3 个效应都是正的, 如果我们只考虑这些主效应, 就可以将此 3 个效应以高水平来实验以使渗透率最大化. 不过, 还需考察任一重要的交互作用. 我们知道, 凡被显著的交互作用牵涉到的主效应, 都没有多大的意义.

图 6.12b 画出了 AC 交互作用和 AD 交互作用的图形. 这些交互作用是求解问题的关键. 由 AC 交互作用看出, 当浓度处于高水平时温度效应很小, 当浓度处于低水平时温度效应很大, 在低浓度、高温度时得到最好的结果. AD 交互作用显示出, 低温时搅拌速度 D 的效应小, 但高温时搅拌速度 D 有大的正的效应. 因此, 当 A 和 D 处于高水平而 C 处于低水平时, 会得到最好的渗透率. 这就允许将甲醛浓度减小至低水平, 这是实验者的另一个目的.

表 6.12　例 6.2 中 2^4 析因实验的因子效应估计及平方和

模型项	效应估计	平方和	贡献百分比
A	21.625	1 870.56	32.639 7
B	3.125	39.062 5	0.681 608
C	9.875	390.062	6.806 26
D	14.625	855.563	14.928 8
AB	0.125	0.062 5	0.001 090 57
AC	−18.125	1 314.06	22.929 3
AD	16.625	1 105.56	19.291 1
BC	2.375	22.562 5	0.393 696
BD	−0.375	0.562 5	0.009 815 15
CD	−1.125	5.062 5	0.088 336 3
ABC	1.875	14.062 5	0.245 379
ABD	4.125	68.062 5	1.187 63
ACD	−1.625	10.562 5	0.184 307
BCD	−2.625	27.562 5	0.480 942
$ABCD$	1.375	7.562 5	0.131 959

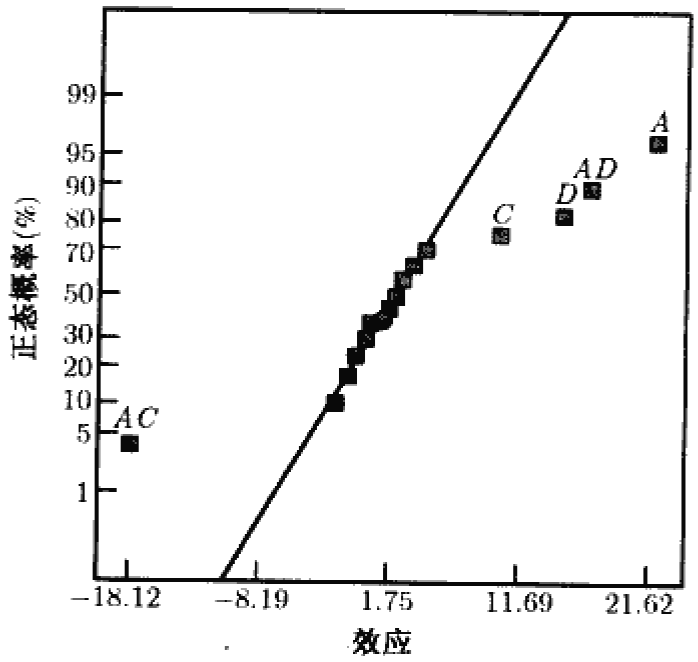
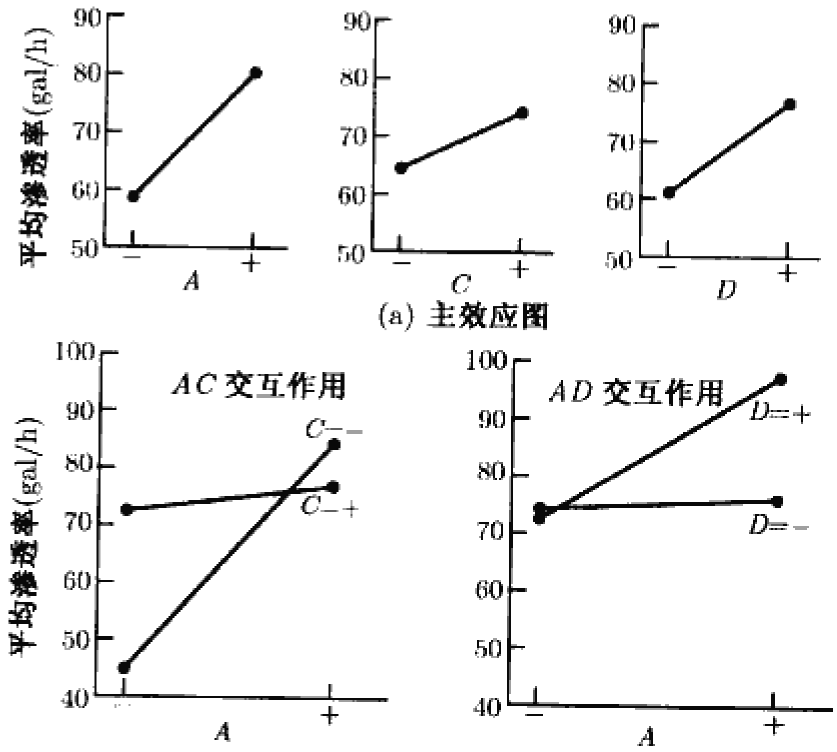


图 6.11　例 6.2 中 2^4 析因实验的效应的正态概率图



(a) 主效应图

(b) 交互作用图

图 6.12　例 6.2 的主效应和交互作用图

1. 设计投影

图 6.11 中的效应可以有另外的解释方式. 因为 B(压强) 不显著且涉及 B 的所有交互作用可被忽略, 所以可以在实验中放弃 B 使得设计变为 A, C, D 的有两次重复的 2^3 析因设计. 考察表 6.10 中的列 A, C, D, 容易看出这些列形成有两次重复的 2^3 设计. 使用这一简化假定, 对数据进行的方差分析概括在表 6.13 中. 由此分析所得的结论与由例 6.2 所得的结论本质上相同. 注意, 使用将 2^4 单次重复投影到有重复的 2^3 中去的方法, 可以得到 ACD 交互作用的估计量以及建立在所谓的**隐藏重复**(hidden replication) 基础上的误差估计量.

表 6.13 小型试验厂渗透率实验中对 A, C, D 的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
A	1 870.56	1	1 870.56	83.36	< 0.000 1
C	390.06	1	390.06	17.38	< 0.000 1
D	855.56	1	855.56	38.13	< 0.000 1
AC	1 314.06	1	1 314.06	58.56	< 0.000 1
AD	1 105.56	1	1 105.56	49.27	< 0.000 1
CD	5.06	1	5.06	< 1	
ACD	10.56	1	10.56	< 1	
误差	179.52	8	22.44		
总和	5 730.94	15			

将无重复的析因设计投影到因子较少的有重复的析因设计中去这一概念非常有用. 一般说来, 如果有一个 2^k 设计的单次重复, 且其中的 $h(h<k)$ 个因子可被忽略因而可被丢失, 则原先的数据对应于留下来的 $k-h$ 个因子的具有 2^h 次重复的二水平析因设计.

2. 诊断检验

将普通的诊断检验应用到 2^k 设计的残差中去. 分析指出, 仅有的显著效应是 $A=21.625, C=9.875, D=14.625, AC=-18.125, AD=16.625$. 如果这一点为真, 则被估计的渗透率由

$$\hat{y}=70.06+\left(\frac{21.625}{2}\right)x_1+\left(\frac{9.875}{2}\right)x_3+\left(\frac{14.625}{2}\right)x_4-\left(\frac{18.125}{2}\right)x_1x_3+\left(\frac{16.625}{2}\right)x_1x_4$$

给出, 其中 70.06 是平均响应, 规范变量 x_1, x_3, x_4 取值为 $+1$ 或 -1. 在试验 (1) 处的预测渗透率是

$$\begin{aligned}\hat{y}=70.06&+\left(\frac{21.625}{2}\right)(-1)+\left(\frac{9.875}{2}\right)(-1)+\left(\frac{14.625}{2}\right)(-1)\\&-\left(\frac{18.125}{2}\right)(-1)(-1)+\left(\frac{16.625}{2}\right)(-1)(-1)=46.22\end{aligned}$$

因为观测值是 45, 所以残差是 $e=y-\hat{y}=45-46.25=-1.25$. 所有 16 个观测值的 y, $\hat{y}$ 和 $e=y-\hat{y}$ 的值如下:

	y	$\hat{y}$	$e=y-\hat{y}$
(1)	45	46.25	−1.25
a	71	69.38	1.63
b	48	46.25	1.75
ab	65	69.38	−4.38
c	68	74.25	−6.25
ac	60	61.13	−1.13
bc	80	74.25	5.75
abc	65	61.13	3.88
d	43	44.25	−1.25
ad	100	100.63	−0.63
bd	45	44.25	0.75
abd	104	100.63	3.38
cd	75	72.25	2.75
acd	86	92.38	−6.38
bcd	70	72.25	−2.25
$abcd$	96	92.38	3.63

残差的正态概率图如图 6.13 所示. 图上的点落在一条直线附近, 它支持了 A, C, D, AC, AD 是仅有的显著效应的论断, 因此满足关于分析的基本假定.

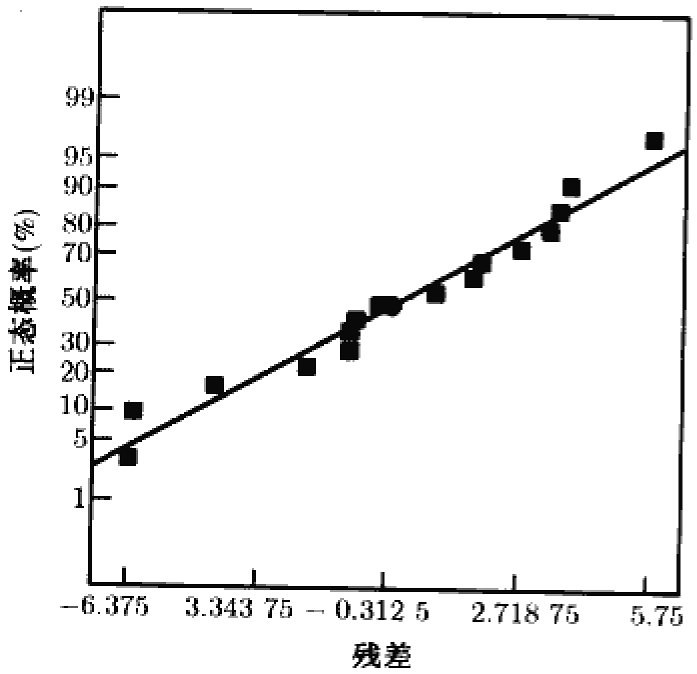


图 6.13　例 6.2 中残差的正态概率图

3. **响应曲面**

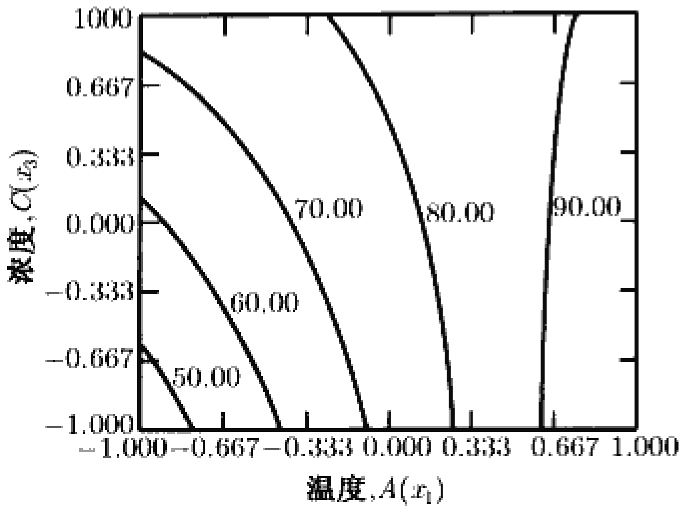
我们用图 6.12 的交互作用图对该实验的结果作一个实际解释. 有时, 我们发现使用响应曲面也有帮助. 响应曲面由回归模型

$$\hat{y}=70.06+4\left(\frac{21.625}{2}\right)x_1+\left(\frac{9.875}{2}\right)x_3+\left(\frac{14.625}{2}\right)x_4-\left(\frac{18.125}{2}\right)x_1x_3+\left(\frac{16.625}{2}\right)x_1x_4$$

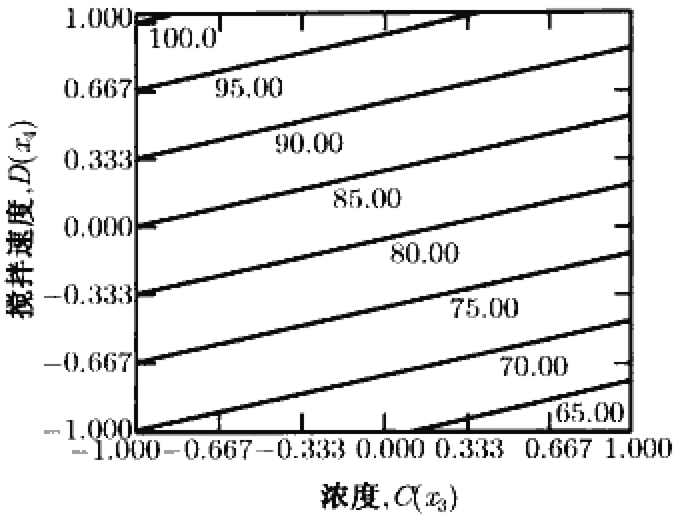
生成. 图 6.14a 显示了搅拌速度在高水平 (即 $x_4=1$) 时的响应曲面等高线图. 将 $x_4=1$ 代入上述模型, 生成等高线, 即

$$\hat{y}=77.372\,5+\left(\frac{38.25}{2}\right)x_1+\left(\frac{9.875}{2}\right)x_3-\left(\frac{18.125}{2}\right)x_1x_3$$

因为模型包含交互作用项, 所以等高线是曲线.



(a) 搅拌速度(D)的等高线图, $x_4=1$



(b) 速度(A)的等高线图, $x_1=1$

图 6.14 例 6.2 渗透率的等高线图

图 6.14b 是温度在高水平 (即 $x_1=1$) 时的响应曲面等高线图. 将 $x_1=1$ 代入回归模型, 得

$$\hat{y} = 80.872\,5 - \left(\frac{8.25}{2}\right)x_3 + \left(\frac{31.25}{2}\right)x_4$$

因为模型仅包含因子 $C(x_3)$ 和 $D(x_4)$, 所以这些等高线是平行直线.

这两张等高线图表明, 如果我们想要极大化渗透率, 变量 $A(x_1)$ 和 $D(x_4)$ 应当是高水平, 过程相对于浓度 C 具有稳健性. 从交互作用图中可以得到类似的结果.

4. 效应的半正态图

也可以使用**半正态图**(half-normal plot) 来代替因子效应的正态概率图. 它是效应估计的绝对值与它们的累积正态概率的关系图. 图 6.15 对于例 6.2 给出了效应的半正态图. 半正态图上的直线始终穿过原点, 而且还应该接近 50%分位点. 许多分析家觉得半正态图易于解释, 特别是在仅有少数效应估计的时候, 比如在实验者进行 8 次试验设计的时候. 一些统计软件包就可以作这些图.

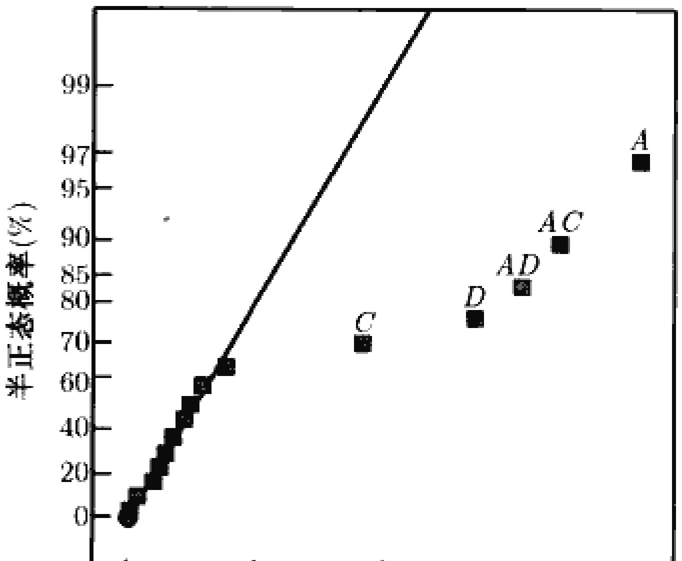


图 6.15 例 6.2 中因子效应的半正态图

5. 分析无重复析因设计的其他方法

关于无重复两水平析因设计的标准分析方法是因子效应估计的正态 (或半正态) 图. 然而, 无重复析因设计在实际中应用中非常广泛, 因此人们提出了许多形式的分析方法, 以克服正态概率图的主观性. Hamada and Balakrishnan(1998) 比较了这些方法中的一部分. 他们发现 Lenth(1989) 提出的方法对于检验效应显著性有很好的功效. 因为它也易于实现, 所以它开始出现在分析无重复析因设计数据的软件包里. 下面给出 Lenth 方法的一个简要介绍.

假定我们感兴趣于 m 个对照, 比如 $c_1, c_2, \cdots, c_m$. 如果设计是无重复 2^k 析因设计, 那么这些对照就对应于 $m = 2^k - 1$ 个因子效应估计. Lenth 方法的基本思想是用最小 (按绝对值) 的对照估计量来估计对照方差. 令

$$s_0 = 1.5 \times \text{中位数}\ (|c_j|)$$

$$PSE = 1.5 \times \ \text{中位数}\ (|c_j| : |c_j| < 2.5 s_0)$$

PSE 称为 "伪标准误", 当没有许多活跃 (显著) 因子时, Lenth 证明它是对照方差的一个合理估计. PSE 用于判断对照的显著性. 一个对照可以与误差边际(margin of error)

$$ME = t_{0.025,d} \times PSE$$

进行比较. 其中自由度定义为 $d = m/3$. 为推导一组对照, Lenth 建议用**误差联合边际**(simultaneous margin of error)

$$SME = t_{\gamma,d} \times PSE$$

其中, t 分布的百分点为 $\gamma = 1 - (1 + 0.95^{1/m})/2$.

为了说明 Lenth 方法, 考虑例 6.2 的 2^4 实验. 用 $s_0 = 1.5 \times |-2.625|$=3.937 5 和 2.5× 3.937 5=9.843 75 进行计算, 得

$$PSE = 1.5 \times |1.75| = 2.625, \quad ME = 2.571 \times 2.625 = 6.75, \quad SME = 5.219 \times 2.625 = 13.70$$

现在, 考虑表 6.12 中的效应估计. SME 准则表明 4 个最大 (指大小) 的效应是显著的, 因为他们的效应估计超过 SME. 根据 ME 准则, 主效应 C 是显著的, 但根据 SME, 它是不显著的. 然而, 由于 AC 的交互作用显然是重要的, 我们在显著效应名单中可能还要包括 C. 该例中, Lenth 方法与我们前面通过检查效应的正态概率图得到的答案是一致的.

一些作者 [见 Hamada and Balakrishnan(1998), Loughin(1998), Loughin and Nobel (1997), 以及 Larntz and Whitcomb(1998)] 已经观测到 Lenth 方法对控制 I 类错误率是失败的, 但模拟的方法可以校正这种方法. Larntz and Whitcomb(1998) 建议用**已调整的乘子**替代原始的 ME 和 SME 乘子如下:

对照数	7	15	31
原始 ME	3.764	2.571	2.218
已调整的 ME	2.295	2.140	2.082
原始 SME	9.008	5.219	4.218
已调整的 SME	4.891	4.163	4.030

这与 Ye and Hamada(2000) 的结果几乎一致.

一般地, Lenth 方法是一个聪明且有用的方法. 然而, 我们推荐它作为通常的效应正态概率图的补充, 而不是替代它.

Bisgaard(1998-1999) 提出了一个名为**条件推断图**(conditional inference chart) 的好的图解方法, 以帮助解释正态概率图. 图解法的目标是帮助实验者判断显著的效应. 当标准差 σ 已知时, 或者它可以从数据中估计时, 这是相对容易的. 在无重复设计中, 没有 σ 的内在估计, 因此, 条件推断图用于帮助实验者估计标准差取值范围的效应大小. Bisgaard 把图建立在下述结果之上: 具有 N 次试验 (对于无重复析因设计, $N = 2^k$) 的两水平设计的效应标准误为

$$\frac{2\sigma}{\sqrt{N}}$$

其中 σ 是单个观测的标准差. ±2 倍的效应的标准误为

$$\pm\frac{4\sigma}{\sqrt{N}}$$

一旦效应被估计, 作出如图 6.16 这样的图, 沿纵坐标 (y 轴) 画出效应估计. 该图已经使用了例 6.2 中的效应估计. 图 6.16 中的横坐标 (即 x 轴) 是一个标准差 (σ) 标量. 这两条直线分别是

$$y = +\frac{4\sigma}{\sqrt{N}}, \quad y = -\frac{4\sigma}{\sqrt{N}}$$

本例中, $N = 16$, 所以直线为 $y = +\sigma$ 和 $y = -\sigma$. 给定标准差 σ 的值, 我们可以把两条直线之间的距离认为是可忽略效应的近似 95%置信区间.

从图 6.16 可以观测到, 如果实验者认为标准差是在 4~8 之间, 那么因子 A, C, D, 以及 AC

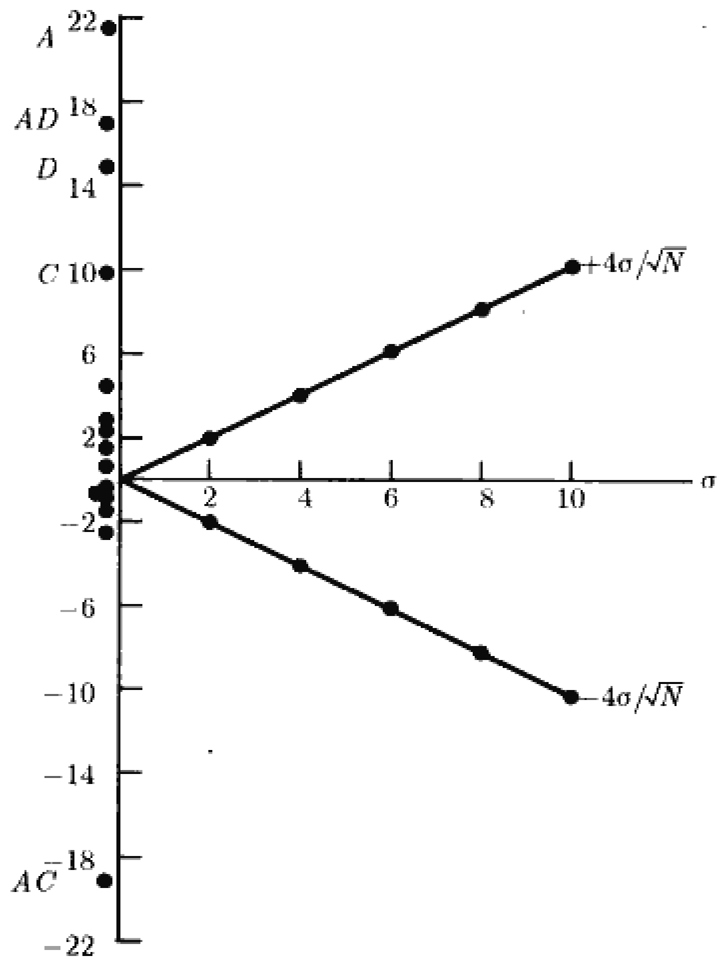


图 6.16　例 6.2 的条件推断图

交互作用和 AD 交互作用是显著的. 如果他或她认为标准差为 10, 因子 C 也许是不显著的. 也就是说, 给定一个 σ 大小的假设, 实验者可以构建一个判断效应近似显著的 "准绳". 该图也可以反过来使用. 例如, 假定我们对因子 C 是否显著不确定. 实验者可以提出问题: $\sigma \geqslant 10$ 是否合理. 如果 $\sigma=10$ 是不合适的, 那么可以得出 C 是显著的结论.

现在给出无重复 2^k 析因设计的 3 个启发性的例子.

例 6.3　析因设计的数据变换

Daniel (1976) 描述了一个 2^4 析因设计, 用来研究钻头的推进速率, 它是 4 个因子 (钻头负荷 (A), 流速 (B), 旋转速度 (C), 泥浆类型 (D)) 的函数. 实验数据如图 6.17 所示.

此实验的效应估计量的正态概率图如图 6.18 所示. 根据此图, 因子 B, C, D 以及 BC 交互作用和 BD 交互作用需要说明. 图 6.19 是残差的正态概率图, 图 6.20 是残差与预测推进速率的关系图, 预测的推进速率来自含有可识别因子的模型. 显然, 关于正态性和等方差性是有问题的. 处理此类问题常用数据变换的方法. 因为响应变量是速率, 所以对数变换是合理的选择.

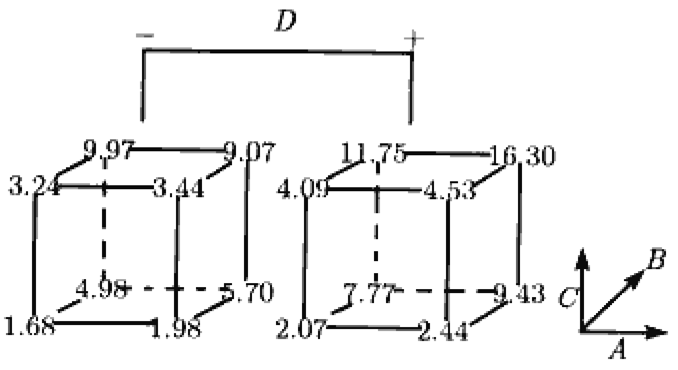


图 6.17　例 6.3 钻头实验中的数据

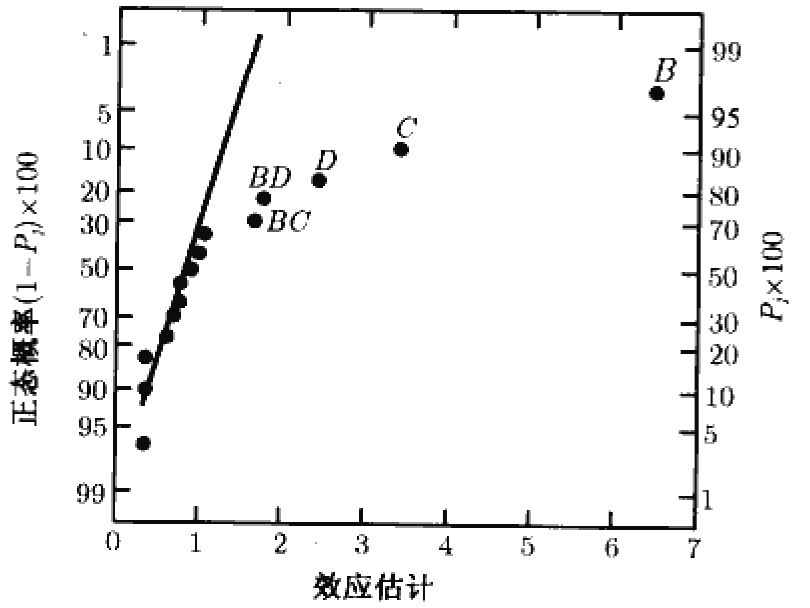


图 6.18　例 6.3 的效应正态概率图

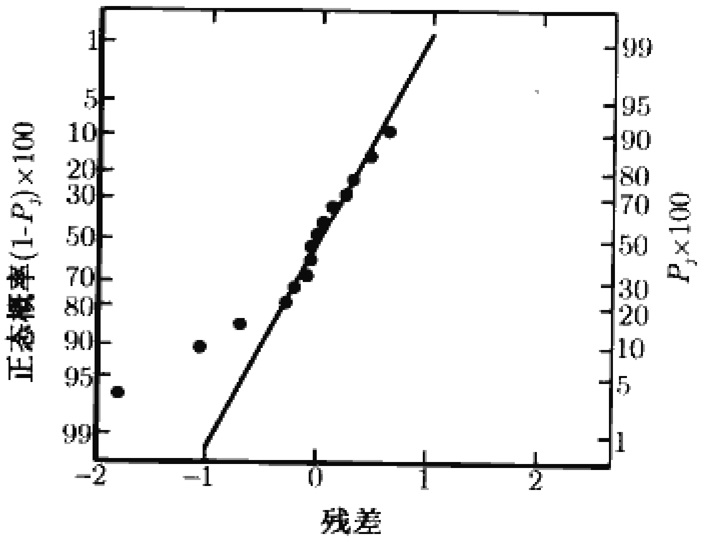


图 6.19　例 6.3 的残差正态概率图

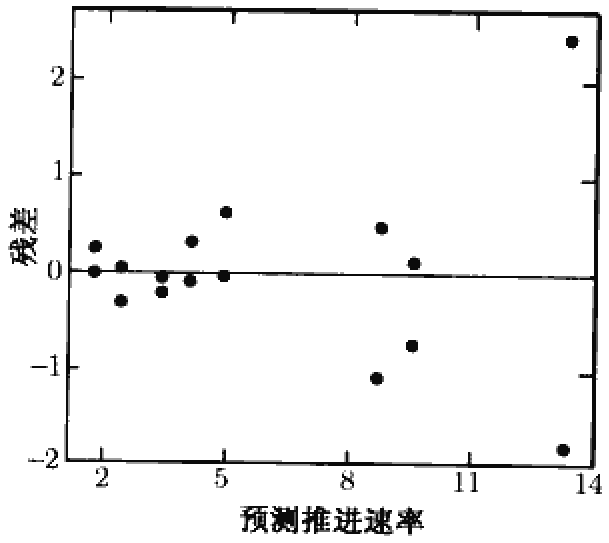


图 6.20　例 6.3 的残差与预测速率的关系图

图 6.21 是在变换 $y^* = \ln y$ 下效应估计量的正态概率图. 我们注意到, 只有因子 B, C, D 起作用, 这样, 解释起来就很简单. 也就是说, 将数据用正确的尺度表示时, 会简化它的结构, 在此说明性模型中那两个交互作用已不再需要了.

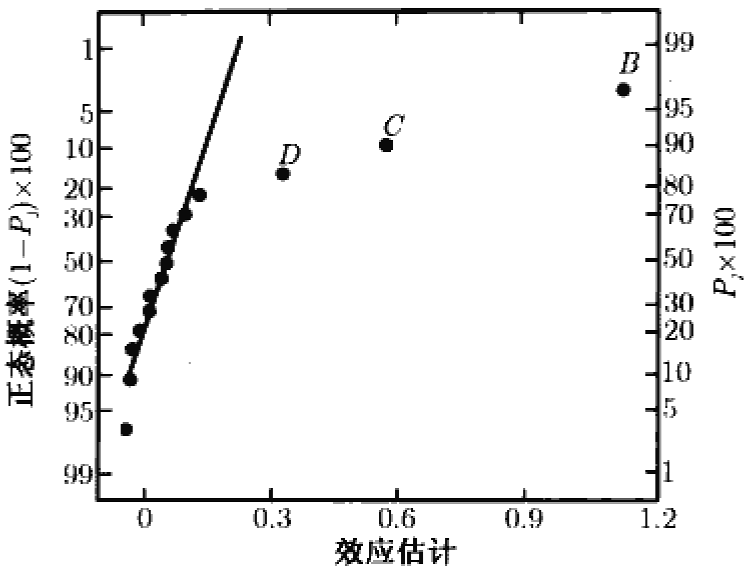


图 6.21 例 6.3 对数变换后的效应与正态概率的关系图

图 6.22 和图 6.23 分别表示残差与正态概率的关系图和残差与预测推进速率的关系图, 这两幅图是在对数意义下. 对含有 B, C, D 的模型而作出. 对现在这些图形终于满意了. 我们的结论是, 对于 $y^* = \ln y$, 模型仅需用因子 B, C, D 就足以说明了. 此模型的方差分析概括在表 6.14 中. 模型的平方和是

$$SS_{模型} = SS_B + SS_C + SS_D = 5.345 + 1.339 + 0.431 = 7.115$$

因此 $R^2 = SS_{模型}/SS_T = 7.115/7.288 = 0.98$, 所以该模型解释了钻头推进速率中大约 98%的变异性.

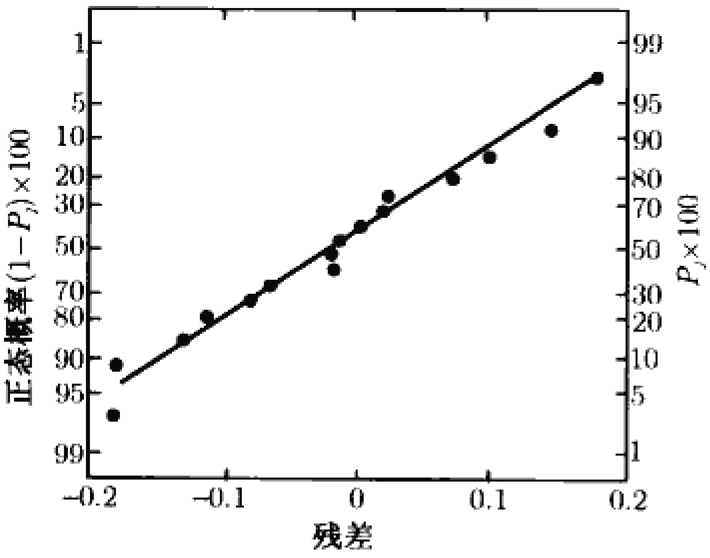


图 6.22 例 6.3 对数变换后的残差与正态概率的关系图

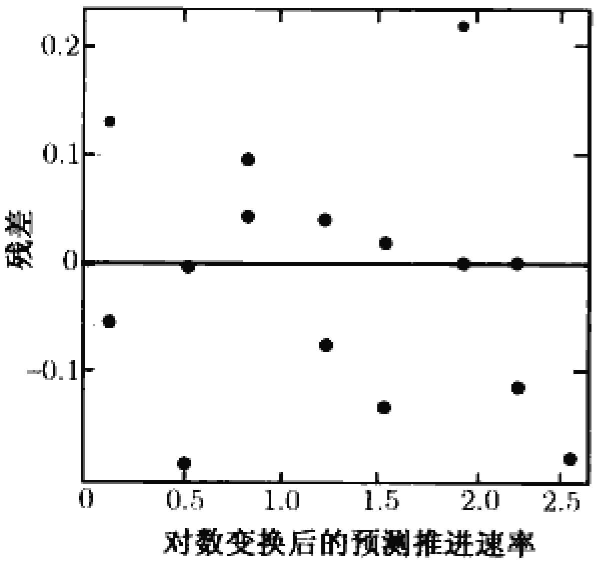


图 6.23 例 6.3 对数变换后的残差与预测速率的关系图

表 6.14 例 6.3 对数变换后的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
B(流速)	5.345	1	5.345	381.79	< 0.000 1
C(旋转速度)	1.339	1	1.339	95.64	< 0.000 1
D(泥浆类型)	0.431	1	0.431	30.79	< 0.000 1
误差	0.173	12	0.014		
总和	7.288	15			

例 6.4 无重复析因设计中的位置效应和分散效应

在一生产商用飞机内壁嵌板和窗嵌板的生产线上做一个 2^4 设计实验. 嵌板是压制成形的, 在当前条件下, 在一压力负荷下每块嵌板的平均疵点数太高 (每块嵌板平均有 5.5 个疵点.) 用 2^4 设计的单次重复来研

究 4 个因子, 实验是在同一压力负荷下进行的. 因子是温度 (A)、模压时间 (B)、树脂流量 (C) 以及压机闭合时间 (D). 此实验的数据如图 6.24 所示.

因子效应的正态概率图如图 6.25 所示. 显然, 两个最大的效应是 $A = 5.75$ 和 $C = -4.25$. 其他因子的效应显得都较小. A 与 C 解释了总变异性的大约 77%, 所以, 结论是, 低的温度 (A) 和高的树脂流量 (C) 会减少嵌板疵点的发生率.

因子	低 (−)	高 (+)
A=温度 (°F)	295	325
B=模压时间 (min)	7	9
C=树脂流量	10	20
D=压机闭合时间 (s)	15	30

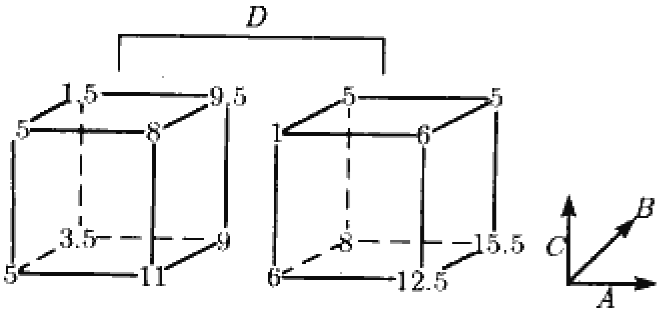


图 6.24　例 6.4 嵌板实验的数据

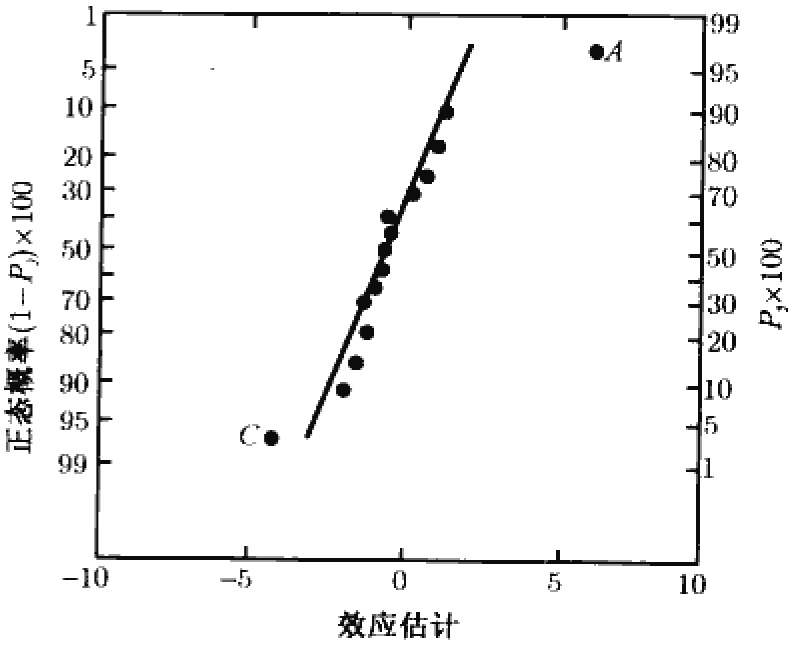


图 6.25　例 6.4 嵌板实验的因子效应的正态概率图

谨慎地进行残差分析是任何实验的一个重要方面. 残差的正态概率图没有显示出异常情况, 但当实验者画出残差与因子 A 到 D 的各个关系图时, 残差与 B(模压时间) 的关系图呈现如图 6.26 所示的模式. 这一因子对于每块嵌板的平均疵点数而论是不重要的, 而在影响生产过程的变异性方面却是十分重要的, 在同一压力负荷下, 少的模压时间会使每块嵌板的平均疵点数有较小的变异性.

从图 6.27 的立方体图可以看出, 模压时间的分散效应也是非常明显的, 图中画出了在由立方体中的因子 A, B, C 确定的点处的每块嵌板平均疵点数及其极差. 当 B 处于高水平时 (图 6.27 中立方体的背面), 平均极差是 $\bar{R}_{B+} = 4.75$, 当 B 处于低水平时, 平均极差是 $\bar{R}_{B-} = 1.25$.

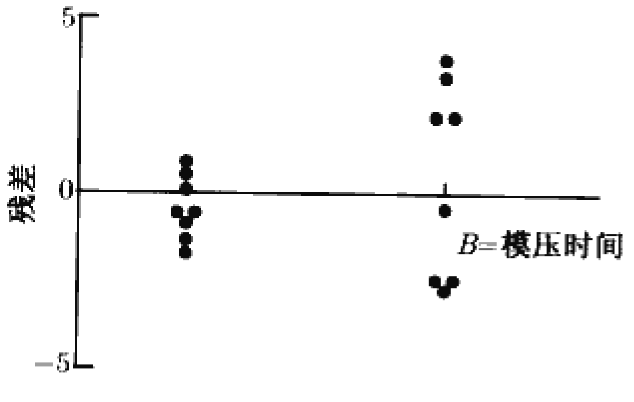


图 6.26　例 6.4 残差与模压时间的关系图

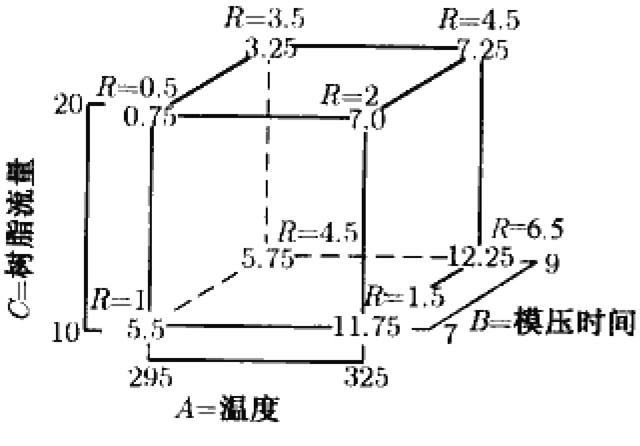


图 6.27　例 6.4 中温度、模压时间和树脂流量的立方体图

这一实验的结果是, 工程师决定, 进行生产时, 取低的温度和高的树脂流量以减少疵点的平均数, 取少的模压时间以减少每块嵌板疵点数的变异性, 并取少的压机闭合时间 (它既对位置和分散性都没有效应). 在一条新的生产线上, 按照这组新的操作条件进行生产, 结果是, 平均每块嵌板少了一个疵点.

2^k 设计的残差能对所研究的问题提供更多的信息. 因为残差可看作为噪音或误差的观测值,

因此, 通过它可对生产过程的变异性作出更深入地了解. 我们可以系统地考察无重复 2^k 设计的残差, 以提供关于生产过程变异性的信息.

考虑图 6.26 的残差图. 当 B 处于低水平时, 8 个残差的标准差是 $S(B^-) = 0.83$; 当 B 处于高水平时, 8 个残差的标准差是 $S(B^+) = 2.72$. 当两个方差 $\sigma^2(B^+)$ 与 $\sigma^2(B^-)$ 相等时, 统计量

$$F_B^* = \ln\frac{S^2(B^+)}{S^2(B^-)} \tag{6.24}$$

近似服从标准正态分布. F_B^* 的值是

$$F_B^* = \ln\frac{S^2(B^+)}{S^2(B^-)} = \ln\frac{(2.72)^2}{(0.83)^2} = 2.37$$

表 6.15 列出了 2^4 设计的完全对照集以及在例 6.4 嵌板实验中的每个残差. 此表的每一列含有相等个数的加号和减号, 因此可对各列的每组符号算出残差的标准差, 即 $S(i^+)$ 和 $S(i^-)$, $i = 1, 2, \cdots, 15$. 于是

$$F_i^* = \ln\frac{S^2(i^+)}{S^2(i^-)} \quad i = 1, 2, \cdots, 15 \tag{6.25}$$

是一统计量, 可用来判断实验**分散效应**的大小. 如果因子 i 为正时试验残差的方差等于因子 i 为负时试验残差的方差, 则 F_i^* 近似服从标准正态分布. F_i^* 的值列于表 6.15 每列的下面.

表 6.15 例 6.4 分散效应的计算

试验	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	$ABCD$	残差
1	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+	−0.94
2	+	−	−	−	−	+	+	−	−	+	+	+	+	−	−	−0.69
3	−	+	−	−	+	−	+	−	+	−	+	+	−	+	−	−2.44
4	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+	−2.69
5	−	−	+	+	−	−	+	−	+	+	−	−	+	+	−	−1.19
6	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+	0.56
7	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+	−0.19
8	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	2.06
9	−	−	+	−	+	+	−	+	−	−	+	−	+	+	−	0.06
10	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+	0.81
11	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+	2.06
12	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−	3.81
13	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	−0.69
14	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	−1.44
15	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	3.31
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−2.44
$S(i^+)$	2.25	2.72	2.21	1.91	1.81	1.80	1.80	2.24	2.05	2.28	1.97	1.93	1.52	2.09	1.61	
$S(i^-)$	1.85	0.83	1.86	2.20	2.24	2.26	2.24	1.55	1.93	1.61	2.11	1.58	2.16	1.89	2.33	
F_i^*	0.39	2.37	0.34	−0.28	−0.43	−0.46	−0.44	0.74	0.12	0.70	−0.14	0.40	−0.70	0.20	−0.74	

图 6.28 是分散效应 F_i^* 的标准正态分布图. 显然, B 是有关生产过程分散性的一个重要因子. 这个方法的进一步讨论见 Box and Meyer(1986) 以及 Myers and Montgomery(2002). 为了使模型的残差完全地转达关于分散效应的信息, 位置模型必须正确地指定. 本章的补充材料提供了更多细节以及一个例子.

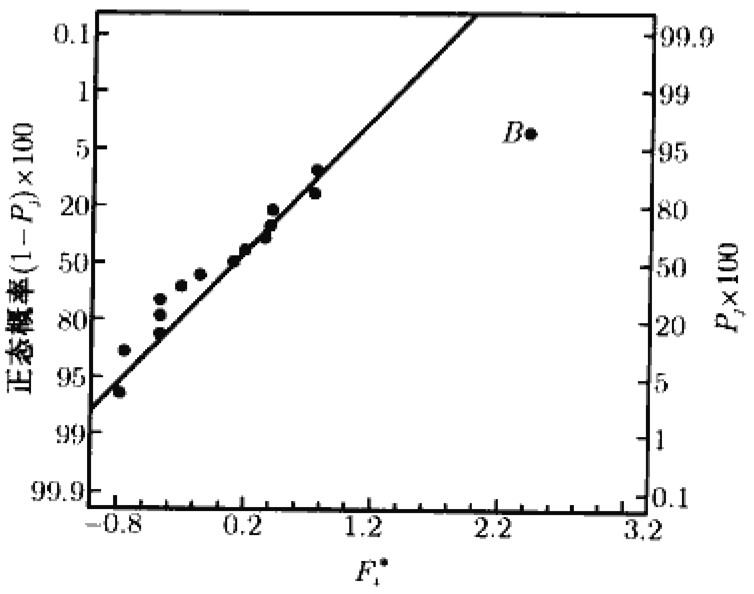


图 6.28 例 6.4 的分散效应 F_i^* 的正态概率图

例 6.5 响应的重复测量

某工程师团队在半导体工厂进行一个立式氧化炉的 2^4 析因设计. 4 个晶片"堆"在炉子里, 响应变量是晶片的氧化物的厚度. 设计因子为温度 (A)、时间 (B)、压强 (C) 和气流 (D). 该实验过程包括: 将 4 个晶片放到炉子里, 设置由实验设计要求的试验条件的过程变量, 处理晶片, 然后测量所有 4 个晶片氧化物厚度. 表 6.16 列出了设计和厚度测量的结果. 该表中, 标为"厚度"的 4 列给出了每个晶片的氧化物厚度测量值, 最后两列为每次试验中 4 个晶片的厚度测量值的样本均值和样本方差.

表 6.16 氧化物厚度实验

标准顺序	试验顺序	A	B	C	D	厚度				$\bar{y}$	s^2
1	10	−1	−1	−1	−1	378	376	379	379	378	2
2	7	1	−1	−1	−1	415	416	416	417	416	0.67
3	3	−1	1	−1	−1	380	379	382	383	381	3.33
4	9	1	1	−1	−1	450	446	449	447	448	3.33
5	6	−1	−1	1	−1	375	371	373	369	372	6.67
6	2	1	−1	1	−1	391	390	388	391	390	2
7	5	−1	1	1	−1	384	385	386	385	385	0.67
8	4	1	1	1	−1	426	433	430	431	430	8.67
9	12	−1	−1	−1	1	381	381	375	383	380	12.00
10	16	1	−1	−1	1	416	420	412	412	415	14.67
11	8	−1	1	−1	1	371	372	371	370	371	0.67
12	1	1	1	−1	1	445	448	443	448	446	6
13	14	−1	−1	1	1	377	377	379	379	378	1.33
14	15	1	−1	1	1	391	391	386	400	392	34
15	11	−1	1	1	1	375	376	376	377	376	0.67
16	13	1	1	1	1	430	430	428	428	429	1.33

该实验的合适分析是将单个晶片厚度测量值视为**重复测量**(duplicate measurement), 而不是重复试验. 如果是真正重复试验, 必须在炉子的一次试验中处理一个晶片. 然而, 由于所有 4 个晶片一起处理, 处理因子 (即, 设计变量的水平) 是同时接到的, 所以晶片厚度测量中的变异性比重复试验中的变异性少得多.

所以, 厚度测量的均值是原先考虑的响应变量的修正值.

表 6.17 列出了实验的效应估计, 用氧化物厚度均值 $\bar{y}$ 作为响应变量. 因子 A 和因子 B 以及交互作用 AB 有大的效应, 总共占平均氧化物厚度变异性的近 90%. 图 6.29 是效应的正态概率图. 通过检查该图, 我们得出结论: 因子 A, B, C 以及交互作用 AB 和 AC 是重要的. 该模型的方差分析表如表 6.18 所示.

表 6.17　例 6.5 的效应估计, 响应变量是平均氧化物厚度

模型项	效应估计	平方和	百分比贡献
A	43.125	7 439.06	67.933 9
B	18.125	1 314.06	12.000 1
C	−10.375	430.562	3.931 92
D	−1.625	10.562 5	0.096 457 3
AB	16.875	1 139.06	10.402
AC	−10.625	451.563	4.123 69
AD	1.125	5.062 5	0.046 231
BC	3.875	60.062 5	0.548 494
BD	−3.875	60.062 5	0.548 494
CD	1.125	5.062 5	0.046 231
ABC	−0.375	0.562 5	0.005 136 78
ABD	2.875	33.062 5	0.301 929
ACD	−0.125	0.062 5	0.000 570 753
BCD	−0.625	1.562 5	0.014 268 8
$ABCD$	0.125	0.062 5	0.000 570 753

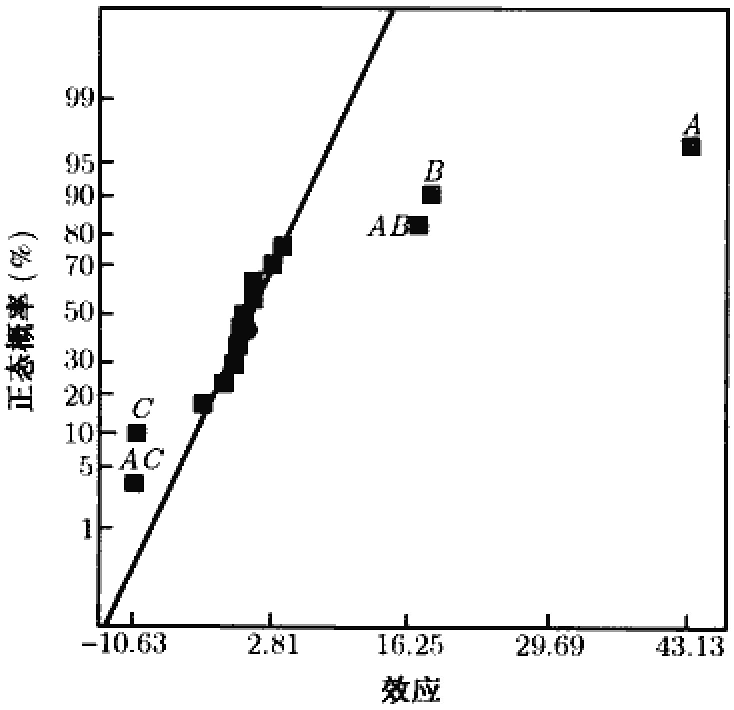


图 6.29　例 6.5 平均厚度响应效应的正态概率图

预测平均氧化物厚度的模型为

$$\hat{y} = 399.19 + 21.56x_1 + 9.06x_2 - 5.19x_3 + 8.44x_1x_2 - 5.31x_1x_3$$

该模型的残差分析是满意的.

表 6.18　例 6.5 平均氧化物厚度的方差分析 (由 Design-Expert)

方差来源	平方和	自由度	均方	F 值	$P>F$
Model	10774.31	5	2154.86	122.35	<0.000
A	7439.06	1	7439.06	422.37	<0.000
B	1314.06	1	1314.06	74.61	<0.000
C	430.56	1	430.56	24.45	0.0006
AB	1139.06	1	1139.06	64.67	<0.000
AC	451.46	1	451.56	25.64	0.0005
Residual	176.12	10	17.61		
Cor Total	10950.44	15			
Std.Dev.	420		R-Squared	0.9839	
Mean	399.19		Adj. R-Squared	0.9759	
C.V.	1.05		Pred.R-Squared	0.9588	
PRESS	450.88		Adeq.Precision	27.967	

因子	系数估计	自由度	标准误	95%CI 低	95%CI 高
Intercept	399.19	1	1.05	396.85	401.53
A-Temp	21.56	1	1.05	19.22	23.90
B-Time	9.06	1	1.05	6.72	11.40
C-Pressure	−5.19	1	1.05	−7.53	−2.85
AB	8.44	1	1.05	6.10	10.78
AC	−5.31	1	1.05	−7.65	−2.97

实验者想要得到平均氧化物厚度为 400Å, 以及产品的规格为厚度在 390Å到 410Å之间. 图 6.30 给出了平均厚度的两幅等高线图: 一个是压强因子 C(即 x_3), 在低水平 (即 $x_3=-1$), 另一个是因子 C(即 x_3) 在高水平 (即 $x_3=+1$). 通过检查这些等高线图, 显然有许多时间和温度 (因子 A 和 B) 的组合能够产生可接受的结果. 但是, 如果压强保持在低水平, 则将 "窗口" 向左移动, 也就是向时间轴的末端移动, 表明为了达到理想的氧化物厚度需要较低的循环时间.

如果我们错误地以为单晶片氧化物厚度测量为重复试验, 则观测得到的结果会很有趣. 表 6.19 给出了基于将实验视为重复 2^4 析因设计的全模型方差分析. 当用平均氧化物厚度作为响应时, 此分析中有许多显著因子, 由此得到的模型远比我们原来得到的模型更复杂. 这是因为表 6.19 中误差方差的估计太小了 ($\hat{\sigma}^2=6.12$). 表 6.19 中的残差均方反映了一次试验内晶片间的变异以及试验间变异的组合. 表 6.18 得到的误差估计比较大, $\hat{\sigma}^2=17.61$, 它主要是试验间变异的度量. 这是用于判断随各次试验而变化的过程变量显著性的最好的误差估计.

一个合乎逻辑的问题是: 若做了表 6.19 中的错误分析, 导致把太多因子识别为重要因子, 这会带来什么不良后果. 答案是, 试图处理或优化不重要因子肯定浪费资源, 而且它必定导致其他感兴趣的响应增加不必要的变异.

当响应有重复测量时, 这些观测几乎总是包含了关于过程变异的某些方面的有用信息. 例如, 如果重复测量是指同一个实验单元用同一个量具进行多次测量, 则它给出了量具能力的一些信息. 如果重复测量在一个实验单元中的不同位置进行, 它们可给出响应变量在整个单元中一致性的一些信息. 本例中, 因为我们对同时处理的 4 个实验单元中的每一个都有一次观测, 因此, 我们得到了过程中试验内变异性的一些信息. 这一信息包含在每次试验的 4 个晶片的氧化物厚度测量的方差中. 确定过程变量是否影响了试验内变异性是有意义的.

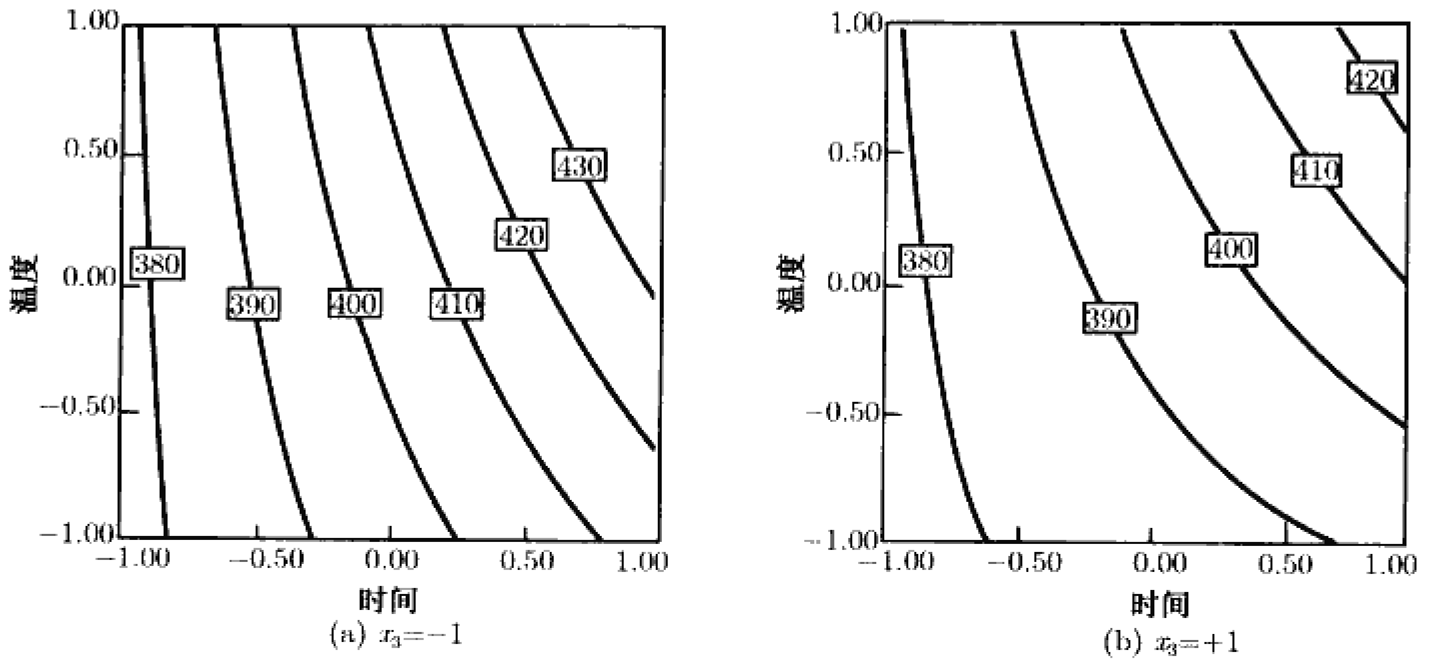


图 6.30 压强 (x_3) 保持常量的平均氧化物的等高线图

表 6.19 单个晶片氧化物厚度响应的方差分析 (由 Design-Expert)

方差来源	平方和	自由度	均 方	F_0	P 值
Model	43801.75	15	2920.12	476.75	<0.0001
A	29756.29	1	29756.29	4858.16	<0.0001
B	5256.25	1	5256.25	858.16	<0.0001
C	1722.25	1	1722.25	281.18	<0.0001
D	42.25	1	42.25	6.90	0.0115
AB	4556.25	1	4556.25	743.88	<0.0001
AC	1806.25	1	1806.25	294.90	<0.0001
AD	20.25	1	20.25	3.31	0.0753
BC	240.25	1	240.25	39.22	<0.0001
BD	240.25	1	240.25	39.22	<0.0001
CD	20.25	1	20.25	3.31	0.0753
ABD	132.25	1	132.25	21.59	<0.0001
ABC	2.25	1	2.25	0.37	0.5473
ACD	0.25	1	0.25	0.041	0.8407
BCD	6.25	1	6.25	1.02	0.3175
ABCD	0.25	1	0.25	0.041	0.8407
Rosidual	294.00	48	6.12		
Lack of Fit	0.000	0			
Pure Error	294.00	48	6.13		
Cor.Total	44095.75	63			

图 6.31 是用 $\ln(s^2)$ 作为响应而得到的效应估计的正态概率图. 回想第 3 章, 我们指出了对数变换一般来讲适用于变异性建模. 没有一个特别强的效应, 而因子 A 和交互作用 BD 是最大的. 如果模型还要包括主效应 B 和 D 以得到分层模型. 那么, $\ln(s^2)$ 的模型是

$$\widehat{\ln(s^2)} = 1.08 + 0.41x_1 - 0.40x_2 + 0.20x_4 - 0.56x_2x_4$$

该模型只解释了 $\ln(s^2)$ 响应中的不足一半的变异性, 作为经验模型, 它做得并不好, 但是得到特别好的方差模型通常是不容易的.

图 6.32 是预测方差 (而不是预测方差的对数) 的等高线图, 其中压强 x_3 在低水平 (这时, 极小化循环时间), 气流 x_4 在高水平. 气流因子的这种选择给出了等高线图区域内的预测方差的最小值.

这里, 实验者要选择设计变量的值, 使得它在过程规范中规定氧化物均值尽可能接近 400Å, 同时使得试验内的变异性较小, 即 $s^2 \leqslant 2$. 用于寻找一组合适的条件的一种可能方法是, 重叠图 6.30 和图 6.32 的等高线图. 图 6.33 显示的就是重叠图, 它用等高线显示了平均氧化物厚度以及约束 $s^2 \leqslant 2$ 的规范. 该图中, 压强保持在低水平, 气流保持在高水平. 图的左上中心的空白区域看成是时间和温度变量的可行域.

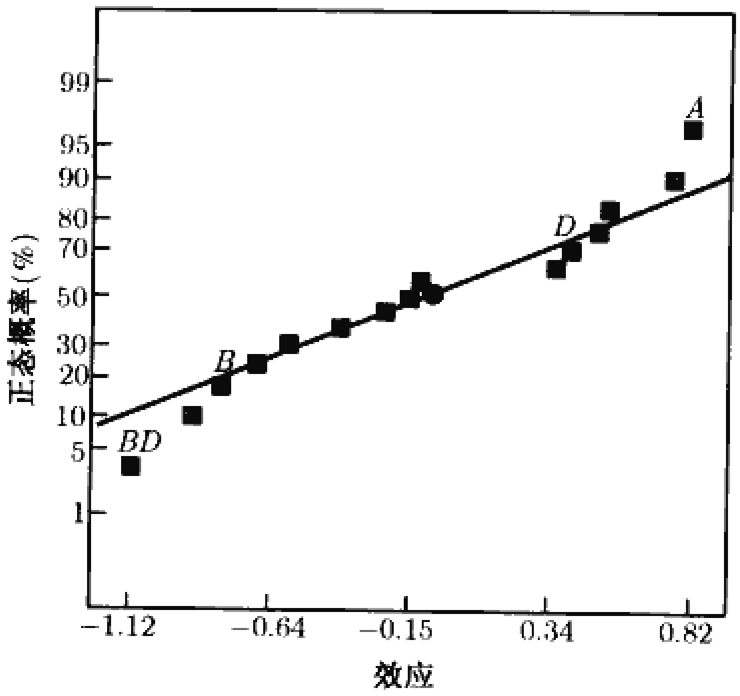


图 6.31　用 $\ln(s^2)$ 作为响应而得到的效应的正态概率图, 例 6.5

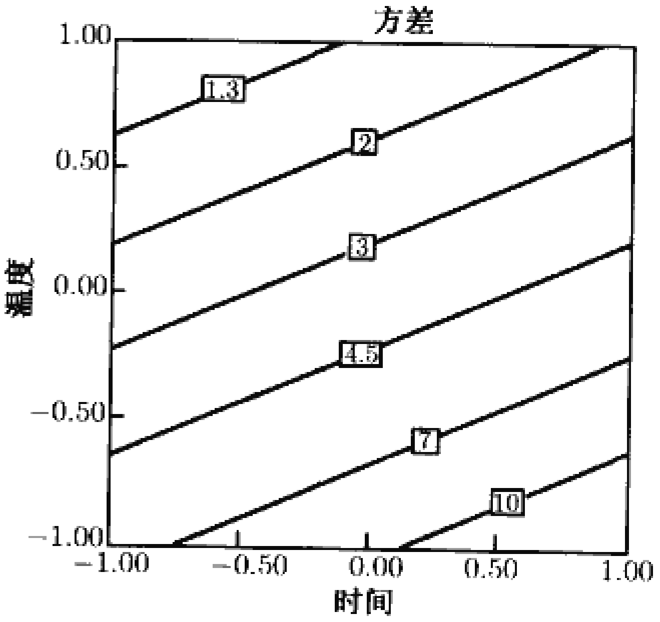


图 6.32　压强在低水平、气流在高水平的 s^2(试验内变异) 等高线图

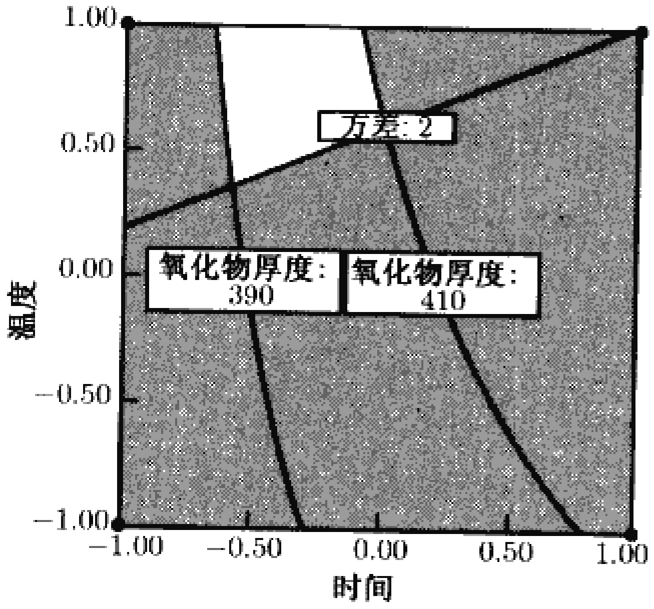


图 6.33　重叠压强在低水平、气流在高水平的平均氧化物厚度及响应 s^2

这是用等高线图同时研究两个响应的简单例子. 我们将在第 11 章中详细讨论.

6.6　附加中心点的 2^k 设计

在利用二水平析因设计时, 一个需要注意的问题是因子效应的线性假定. 当然, 完全的线性是不必要的, 甚至当线性假定仅仅相当近似地成立, 2^k 系统也工作得很好. 事实上, 我们注意到,

如果在主效应即一阶模型中增加**交互作用项**(interaction term), 得

$$y=\beta_0+\sum_{j=1}^{k}\beta_j x_j+\sum_{i<j}\sum\beta_{ij}x_i x_j+\varepsilon \tag{6.26}$$

于是我们就有了可以表示响应函数的一些弯曲性的模型. 当然, 这些弯曲性是由交互作用项 $\beta_{ij}x_i x_j$ 引起的平面扭曲得到的.

有时, (6.26) 式不足以模拟响应函数的弯曲性. 在这种情况下, 要考虑的一个合理的模型是

$$y=\beta_0+\sum_{j=1}^{k}\beta_j x_j+\sum_{i<j}\sum\beta_{ij}x_i x_j+\sum_{j=1}^{k}\beta_{jj}x_j^2+\varepsilon \tag{6.27}$$

其中, β_{jj} 为纯二阶效应即**二次效应**(quadratic effect). (6.27) 式称为**二阶响应曲面模型**.

在二水平析因实验中, 我们通常期望拟合 (6.26) 式的一阶模型, 但我们必须注意 (6.27) 式的二阶模型更合适的可能性. 有一种方法, 在 2^k 析因实验中, 重复某些点的方法将提供针对来自二阶效应的弯曲性的保护, 并可得到一个独立的误差估计. 这一方法由在 2^k 设计中加进一**中心点**而构成. 在这点 $x_i=0\,(i=1,\,2,\,\cdots,\,k)$ 处做 n 次重复试验. 在设计中心处加进重复试验的一条重要的理由是, 中心点不影响 2^k 设计中通常的效应估计量. 我们假定 k 个因子是**定量**的.

为说明这一方法, 考虑在每个因子点 $(-,\,-)$、$(+,-)$、$(-,+)$ 和 $(+,\,+)$ 处有一个观测值而在中心点 (0, 0) 处有 n_C 个观测值的 2^2 设计. 图 6.34 和图 6.35 说明了这一点.

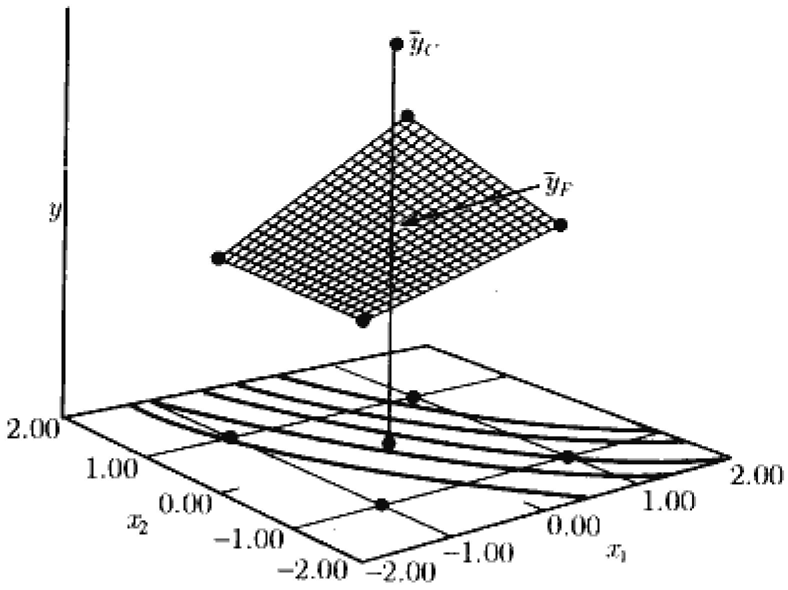


图 6.34 有中心点的 2^2 设计

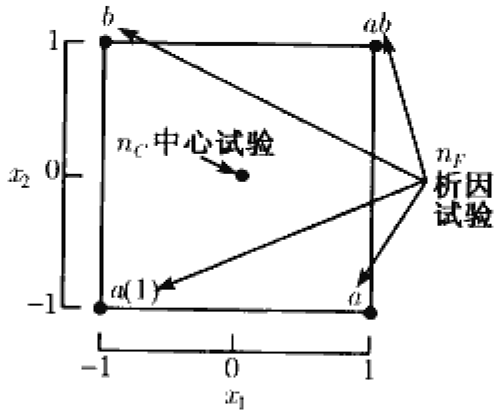


图 6.35 有中心点的 2^2 设计

设 $\bar{y}_F$ 是在 4 个因子点处 4 次试验的平均值, $\bar{y}_C$ 是中心点处 n_C 次试验的平均值. 如果 $\bar{y}_F-\bar{y}_C$ 的差很小, 则中心点就在或者靠近通过因子点的平面上, 因此没有弯曲. 而如果 $\bar{y}_F-\bar{y}_C$ 较大, 则出现弯曲. 单自由度的**纯二次项弯曲部分的平方和**为

$$SS_{\text{纯二次项}}=\frac{n_F n_C(\bar{y}_F-\bar{y}_C)^2}{n_F+n_C} \tag{6.28}$$

其中, 一般说来, n_F 是析因设计的点的个数. 平方和也许合并在方差分析中, 此量与误差均方相比, 可用来对弯曲性进行检验. 更具体地, 当加进的点是 2^k 设计的中心时, 则 [由 (6.28) 式] 弯曲性检验实际上是检验假设

$$H_0:\sum_{j=1}^{k}\beta_{jj}=0,\qquad H_1:\sum_{j=1}^{k}\beta_{jj}\neq 0$$

而且，如果设计的因子点是无重复的，则可以用中心点处的 n_C 个观测值来构造一个自由度为 n_C-1 的误差估计量. 详见本章的**补充材料**.

例 6.6 我们通过再考虑例 6.2 的小规模试验性工厂实验，来说明附加中心点的 2^k 设计. 这是一个无重复 2^4 设计. 见表 6.10 的原始实验. 假定在该实验中增加 4 个中心点，在 4 个点 $x_1=x_2=x_3=x_4=0$ 的渗透率为 73, 75, 66, 69. 4 个中心点的均值为 $\bar{y}_C=70.75$, 16 个析因试验的均值为 $\bar{y}_F=70.06$. 因为 $\bar{y}_C$ 和 $\bar{y}_F$ 非常接近，我们怀疑不存在强烈的弯曲.

表 6.20 概括出此实验的方差分析. 在表的上部，我们拟合了全模型. 利用中心点算得的均方误差如下：

$$MS_E=\frac{SS_E}{n_C-1}=\frac{\sum\limits_{\text{中心点}}(y_i-\bar{y}_C)^2}{n_C-1}\tag{6.29}$$

因此，在表 6.20 中，

$$MS_E=\frac{\sum\limits_{i=1}^{4}(y_i-70.75)^2}{4-1}=\frac{48.75}{3}=16.25$$

由 (6.28) 式，用差 $\bar{y}_F-\bar{y}_C=70.06-70.75=-0.69$ 计算方差分析表中的纯二次项 (弯曲) 平方和，算得如下：

$$SS_{\text{纯二次项}}=\frac{n_Fn_C(\bar{y}_F-\bar{y}_C)^2}{n_F+n_C}=\frac{(16)(4)(-0.69)^2}{16+4}=1.51$$

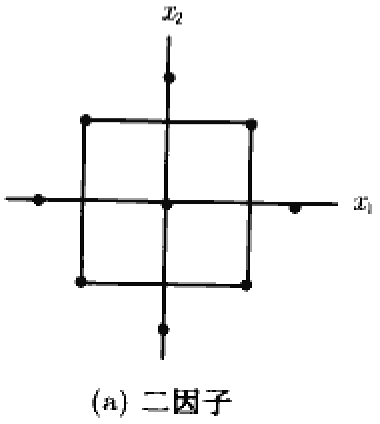
方差分析说明在所考察的范围内没有证据表明响应的二阶弯曲部分存在. 也就是说，不能拒绝零假设 $H_0:\beta_{11}+\beta_{22}+\beta_{33}+\beta_{44}=0$. A, C, D, AC 和 AD 为显著效应. 简化模型的方差分析列在表 6.20 的下部. 这个分析的结果与例 6.2 一致，那里是用正态概率图法分离出重要效应.

在例 6.6 中，我们认为没有迹象表明二次项效应存在. 即，含有 A, C, D, 以及交互作用 AC 和 AD 的一阶模型是合适的. 但是也存在二次项 (x_i^2) 的情况. 例如，对于 $k=2$ 个设计因子的情形. 若弯曲性检验是显著的，则只能假定二阶模型，如

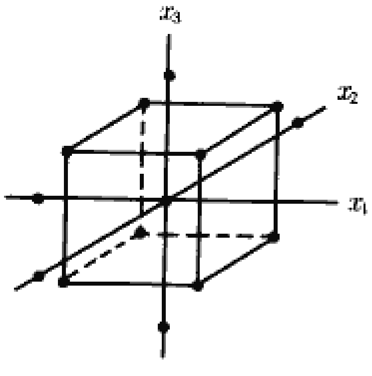
$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_{12}x_1x_2+\beta_{11}x_1^2+\beta_{22}x_2^2+\varepsilon$$

不幸的是，我们不能在该模型中估计未知参数 (β)，因为有 6 个参数要估计，图 6.35 的加上中心点的 2^2 设计只有 5 个独立试验.

解决这个问题的简单而高效的方法是在 2^k 设计中增加 4 个**轴试验**(axial run)，如图 6.36a 所示的 $k=2$ 情形. 得到的设计，称为**中心复合设计**(central composite design)，可以用于拟合二阶模型. 图 6.36b 显示了 $k=3$ 个因子的中心复合设计. 该设计有 $14+n_C$ 试验 (通常 $3\leqslant n_C\leqslant 5$)，能够有效地拟合 $k=3$ 个因子情形下的含有二阶模型 10 个参数的.



(a) 二因子



(b) 三因子

图 6.36 中心复合设计

中心复合设计广泛用于建立二阶响应曲面模型. 第 11 章中将详细讨论这些设计.

表 6.20 例 6.6 的方差分析

全模型的方差分析					
方差来源	平方和	自由度	均方	*F* 值	*P*>*F*
Model	5 730.94	15	382.06	23.51	0.012 1
A	1 870.56	1	1 870.56	115.11	0.001 7
B	39.06	1	39.06	2.40	0.218 8
C	390.06	1	390.06	24.00	0.016 3
D	855.56	1	855.56	52.65	0.005 4
AB	0.063	1	0.063	3.846E-003	0.954 4
AC	1 314.06	1	1 314.06	80.87	0.002 9
AD	1 105.56	1	1 105.56	68.03	0.003 7
BC	22.56	1	22.56	1.39	0.323 6
BD	0.56	1	0.56	0.035	0.864 3
CD	5.06	1	5.06	0.31	0.615 7
ABC	14.06	1	14.06	0.87	0.420 9
ABD	68.06	1	68.06	4.19	0.133 2
ACD	10.56	1	10.56	0.65	0.479 1
BCD	27.56	1	27.56	1.70	0.283 8
ABCD	7.56	1	7.56	0.47	0.544 1
Pure Quadratic					
Curvature	1.51	1	1.51	0.093	0.780 2
Pure error	48.75	3	16.25		
Cor Total	5 781.20	19			
简约模型的方差分析					
方差来源	平方和	自由度	均方	*F* 值	*P*>*F*
Model	5 535.81	5	1 107.16	59.02	<0.000
A	1 870.56	1	1 870.56	99.71	<0.000
C	390.06	1	390.06	20.79	0.000 5
D	855.56	1	855.56	45.61	<0.000
AC	1 314.06	1	1 314.06	70.05	<0.000
AD	1 105.56	1	1 105.56	58.93	< 0.000
Pure Quadratic					
Curvature	1.51	1	1.51	0.081	0.780 9
Residual	243.87	13	18.76		
Lack of Fit	195.12	10	19.51	1.20	0.494 2
Pure error	48.75	3	16.25		
Cor Total	5 781.20	19			

结束本节前, 我们提出几个有用的建议和关于中心点使用的意见.

(1) 当析因实验在一个正在运行的过程中进行时, 考虑使用目前的运行条件 (或配方) 作为设计的中心点. 这通常确保实验中的部分操作人员能在熟悉的条件下完成任务, 所以得到的结果 (至少是这些试验) 不会比现在的结论差.

(2) 当析因实验中的中心点对应于通常的运行配方时, 实验者可以利用中心点的响应观测值

来粗略检查实验中是否有任何"不寻常的"现象. 也就是, 中心点的响应必须和常规过程运行的响应观测类似. 通常操作人员持有监控过程性能的控制图. 有时, 中心点响应直接画在控制图上作为检查实验中正运行过程的方法.

(3) 考虑以非随机次序在中心点进行重复. 具体来说, 在实验开始时或开始之初, 在中心点做一次或两次试验, 实验中期做一两次, 实验结束前, 做一两次. 随着中心点试验逐步进行, 实验者可以在实验中对过程稳定性有一个粗略的检查. 例如, 实验进行中, 响应有一个趋势, 作中心点响应与时间顺序的关系图可以说明这一点.

(4) 有时实验必须在很少或没有关于过程变异性先验信息的条件下进行. 这种情况下, 做 2 ~ 3 个中心点试验作为实验中的最初几个试验是有帮助的. 这些试验可以提供变异性的初步估计. 如果变异性的大小看起来是合理的, 则继续试验; 如果观测到大得超过预期的 (或合理的) 变异性, 则停止试验. 在其余实验继续进行下去之前, 研究变异性为何如此之大是大有帮助的.

(5) 通常, 当所有设计因子是定量的时候, 我们可以使用中心点. 但是, 有时会有一个或多个定性或分类变量以及几个定量因子. 这时, 仍然可以使用中心点. 例如, 考虑一个实验, 有两个定量因子 (时间和温度) 每个因子有两个水平, 还有一个定性因子 (晶体类型), 也有两个水平 (有机的和无机的). 图 6.37 显示了这些因子的 2^3 设计. 注意到中心点被放在包含定性因子的立方体的对面. 换句话说, 只要这些子空间仅包含定量因子, 中心点就可以在这些定性因子的高水平和低水平的处理组合上进行试验.

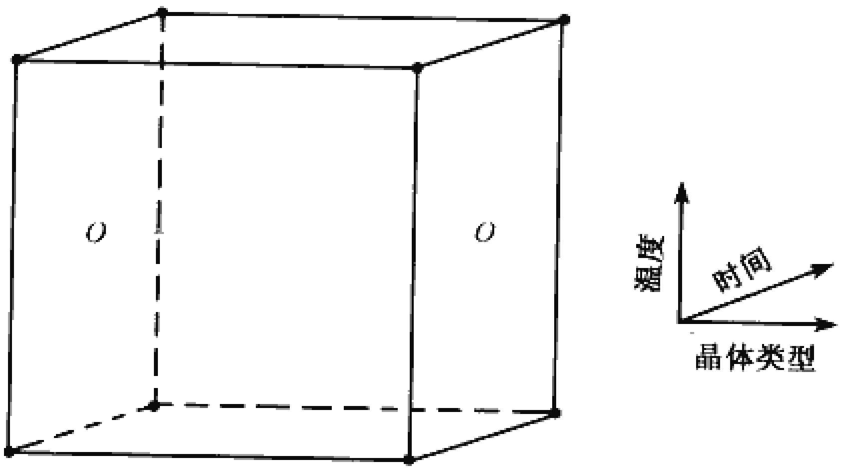


图 6.37 含有一定性因子和中心点的 2^3 析因设计

6.7 使用规范化设计变量的理由

读者已注意到, 在本章中, 我们在进行所有 2^k 析因设计和模型拟合时都采用了**规范设计变量**, $-1 \leqslant x_i \leqslant +1$, 而不是来用按它们**原始**单位 (有时称实际、自然或工程单位) 的设计因子. 若采用工程单位, 与规范单位分析相比, 我们将得到不同的数值结果, 通常这个结果不容易解释.

为了说明两个分析间的一些差异, 考虑下面的实验. 一个简单 DC 电路连接两个不同的电阻, 1Ω 和 2Ω. 电路也包含一只电流表和一个可变输出电源. 当电阻安装在电路中时, 调整电源直到电流达到 4 安培或 6 安培. 用伏特计读取电源的电压输出. 进行 2^2 析因设计的两次重复, 表 6.21 给出了结果. 我们知道欧姆定律确定了与测量误差无关的电压观测. 无论如何, 对由经验模型得到的数据的分析给出了一些关于所设计的实验中规范单位和工程单位的值的信息.

表 6.21 电路实验

I(安培)	R(欧姆)	x_1	x_2	V(伏特)
4	1	−1	−1	3.802
4	1	−1	−1	4.013
6	1	1	−1	6.065
6	1	1	−1	5.992
4	2	−1	1	7.934
4	2	−1	1	8.159
6	2	1	1	11.865
6	2	1	1	12.138

表 6.22 和表 6.23 分别给出了用通常的规范变量 (x_1 和 x_2) 以及工程单位作为设计变量的回归模型. Minitab 用于进行计算. 首先考虑表 6.22 中的规范变量分析. 设计是正交的, 规范变

表 6.22 用规范变量的电路实验的回归分析

```
The regression equation is
V  =  7.50  +  1.52 x1  +  2.53 x2  +  0.458 x1x2
Predictor        Coef         StDev          T          P
Constant       7.49600       0.05229      143.35      0.000
x1             1.51900       0.05229       29.05      0.000
x2             2.52800       0.05229       18.34      0.000
x1x2           0.45850       0.05229        8.77      0.001
S=0.1479          R-Sq=99.9%            R-Sq(adj)=99.8%
Analysis of Variance
Source             DF        SS          MS          F          P
Regression          3      71.267      23.756     1085.95     0.000
Residual Error      4       0.088       0.022
Total               7      71.354
```

表 6.23 用工程单元的电路实验的回归分析

```
The regression equation is
V  =  −0.806  +  0.144  I  +  0.471  R  +  0.917  IR
Predictor        Coef         StDev          T          P
Constant      −0.8055        0.8432       −0.96      0.394
I              0.1435        0.1654        0.87      0.434
R              0.4710        0.5333        0.88      0.427
IR             0.9170        0.1046        8.77      0.001
S=0.1479          R-Sq=99.9%            R-Sq(adj)=99.8%
Analysis of Variance
Source             DF        SS          MS          F          P
Regression          3      71.267      23.756     1085.95     0.000
Residual Error      4       0.088       0.022
Total               7      71.354
```

量也是正交的. 注意到所有主效应 (x_1= 电流) 和 (x_2= 电阻器) 都是显著的. 交互作用也很显著. 在规范变量分析中, 模型系数的大小是直接比较的, 即, 它们都是无尺度的, 它们测量了效应在一个单位区间的变化. 而且, 它们估计的精度都是相同的 (注意到所有 3 个系数的标准误都是 0.053). 交互作用效应小于任一个主效应, 电流的效应稍大于电阻效应的一半. 这暗示着, 在因子研究的范围内, 电阻是一个更为重要的变量. 对于确定因子效应的相对大小, 规范变量非常有效.

现在考虑基于工程单位的分析, 如表 6.23 所示. 在该模型中, 仅有交互作用是显著的. 交互作用项的模型系数为 0.917 0, 标准误为 0.104 6. 可以构建检验交互作用为 1 的假设的 t 统计量:

$$t_0 = \frac{\hat{\beta}_{IR} - 1}{se(\hat{\beta}_{IR})} = \frac{0.917\,0 - 1}{0.104\,6} = -0.793\,5$$

该检验统计量的 P 值是 P=0.76. 因此, 我们不能拒绝系数为 1 的零假设, 这与欧姆定律一致. 注意到回归系数是有量纲的, 它们估计的精度并不不同. 这是因为用工程单位因子的实验设计不是正交的.

因为截距项和主效应是不显著的, 我们可以考虑仅用交互作用项 IR 拟合模型. 结果如表 6.24 所示. 交互作用项回归系数的估计和前面工程单位的估计是不同的, 因为工程单位的设计不是正交的. 系数也为 1.

表 6.24　电路实验的回归分析 (仅考虑交互作用项)

```
The regression equation is
V  =  1.00 IR
Predictor           Coef          StDev          T           P
Noconstant
IR               1.00073        0.00550       181.81       0.000
S=0.1255
Analysis of Variance
Source          DF        SS          MS            F           P
Regression       3    71.267      23.756      1085.95       0.000
Residual Error   4     0.088       0.022
Total            7    71.354
```

一般地, 工程单位不能直接比较, 但是它们有物理含义, 正如在当前的例子中的一样. 这可能得到基于内在机理的简化. 在大多数情况中, 最好还是选择规范单位分析. 它通常不会发生基于内在机理的简化 (如同在我们的例子中的). 在实践中, 规范变量可以帮助实验者认清设计因子的相对重要性.

6.8　思　考　题

*6.1　工程师研究切割速度 (A)、工具的几何形状 (B) 以及切割角 (C) 对于切削工具寿命 (以小时计) 的效应. 每一因子选取两个水平, 采用有 3 次重复的 2^3 析因设计. 结果如下:

(a) 估计因子效应. 哪些效应较大?

(b) 用方差分析来确认你对 (a) 的结论.

(c) 根据实验结果, 写出预测寿命 (单位: 小时) 的回归模型.

(d) 分析残差. 有明显的问题吗?

(e) 基于主效应和交互作用图的分析, 你推荐 A, B, C 的哪一种规范因子水平?

A	B	C	处理组合	重 复		
				I	II	III
−	−	−	(1)	22	31	25
+	−	−	a	32	43	29
−	+	−	b	35	34	50
+	+	−	ab	55	47	46
−	−	+	c	44	45	38
+	−	+	ac	40	37	36
−	+	+	bc	60	50	54
+	+	+	abc	39	41	47

*6.2 重新考虑思考题 6.1 中的 (c), 用回归模型来生成工具寿命的响应曲面和等高线图. 解释这些图形. 在这个过程中它们提供了关于理想操作条件的信息吗?

6.3 求思考题 6.1 中因子效应的标准误和近似 95%置信区间. 此分析的结果与由方差分析得到的结论一致吗?

6.4 取适当尺度的 t 分布, 画出思考题 6.1 中因子效应的图形. 从图中能恰当地识别出重要的因子吗? 将由图中所得的结论与由方差分析所得的结论进行比较.

*6.5 有一台用来在印刷电路板上刻痕的刻纹机. 刻划时在电路板表面上的摆动水平可以看作为刻痕空间位置摆动的主要来源. 考虑影响摆动的两个因子: 刀刃尺寸 (A) 和刻划速度 (B). 选择两种刀刃尺寸 (1/16 英寸和 1/8 英寸) 和两种速度 (40 转/分和 90 转/分). 在每组条件下刻 4 块电路板, 结果如下. 响应变量是摆动, 在每块试验电路板上用 3 只传感器 (x, y, z) 的合成量来度量它.

A	B	处理组合	重 复			
			I	II	III	IV
−	−	(1)	18.2	18.9	12.9	14.4
+	−	a	27.2	24.0	22.4	22.5
−	+	b	15.9	14.5	15.1	14.2
+	+	ab	41.0	43.9	36.3	39.9

(a) 分析此实验的数据.

(b) 画出残差的正态概率图, 并画出残差与预测摆动水平的关系图. 解释这些图形.

(c) 画出 AB 交互作用图. 解释此图. 你推荐刀刃尺寸和刻划速度的哪一水平用于日常生产?

*6.6 重新考虑思考题 6.1, 假设实验者仅在重复 I 中进行了 8 次试验. 另外, 他设定了 4 个中心点, 得到了以下响应数据: 36, 40, 43, 45.

(a) 估计因子效应, 哪个效应最大?

(b) 进行方差分析, 包括纯二次弯曲性的检查, 你的结论是什么?

(c) 根据实验结果, 写出一个预测工具寿命的合适模型. 这个模型和思考题 6.1 中 (c) 部分的模型有何不同?

(d) 分析残差.

(e) 关于这个过程的合适运行条件, 你得到了什么结论?

6.7 进行一个用以改善化学过程产率的实验. 选取 4 个因子, 采用有两次重复的完全随机化实验. 结果如下表所示:

处理组合	重复 I	重复 II	处理组合	重复 I	重复 II
(1)	90	93	d	98	95
a	74	78	ad	72	76
b	81	85	bd	87	83
ab	83	80	abd	85	86
c	77	78	cd	99	90
ac	81	80	acd	79	75
bc	88	82	bcd	87	84
abc	73	70	$abcd$	80	80

(a) 估计因子效应.

(b) 写出方差分析表, 确定哪些因子对产率是重要的.

(c) 假设所有 4 个因子的变化范围都是从 -1 到 $+1$(规范变量), 写出一个预测产率的回归模型.

(d) 在正态概率纸上画出残差与预测产率的关系图. 残差分析令人满意吗?

(e) 两个三因子交互作用 ABC 和 ABD 显示出有较大的效应. 对因子 A, B, C 画一立方体图, 并在每个角上标出相应的平均产率. 对因子 A, B, D 也同样画一张. 这两张图对数据解释有帮助吗? 进行生产时, 关于这 4 个变量你会推荐在哪一位置上的呢?

6.8 一位细菌学家考察两种不同的培养基和两种不同的滋长时间对于某病毒的效应. 她采用有 6 次重复的 2^2 设计, 以随机顺序做试验. 分析下面的病毒滋长数据, 并写出恰当的结论. 分析残差并论述模型的适合性.

时间 (小时)	培养基 1		培养基 2	
12	21	22	25	26
	23	28	24	25
	20	26	29	27
18	37	39	31	34
	38	38	29	33
	35	36	30	35

6.9 一位主管生产饮料瓶的工程师考虑两种不同的 32 盎司瓶子在运送过程中所需的时间, 一种是玻璃的, 另一种是塑料的, 12 瓶一箱. 两位工人来执行下述任务: 将 40 箱产品用标准手推车运至距原地 50 英尺处, 并堆放整齐. 采用有 4 次重复的 2^2 析因设计. 观测到的时间如下表所示. 分析这些数据并写出恰当的结论. 分析残差并论述模型的适合性.

瓶子类型	工人 1		工人 2	
玻璃	5.12	4.89	6.65	6.24
	4.98	5.00	5.49	5.55
塑料	4.95	4.43	5.28	4.91
	4.27	4.25	4.75	4.71

6.10 在思考题 6.9 中, 工程师也感兴趣于由于瓶子类型的不同而导致工人的潜在的疲劳差. 作为所需要

的劳动量的量度，他测量了由于工作引起的心率（脉搏）的升高，结果如下．分析这些数据并写出结论．分析残差并论述模型的适合性．

瓶子类型	工人			
	1		2	
玻璃	39	45	20	13
	58	35	16	11
塑料	44	35	13	10
	42	21	16	15

6.11 计算思考题 6.10 中因子效应的近似 95% 置信区间．这个分析的结果和思考题 6.10 中方差分析的结果一致吗？

6.12 在*AT & Technical Journal* (March/April 1986, Vol. 65, pp. 39-50) 上的一篇论文描述了将二水平析因设计应用于集成电路生产的例子．一个基本的生产步骤是，在已抛光的硅片上生成一层外延层．硅片固定在基座上并定位于一个钟形烧结炉内，然后引入化学蒸气．基座不断地旋转和受热，直到外延层足够厚为止．进行试验时使用两个因子：砷流速率 (A) 和沉积时间 (B)．进行 4 次重复，测量外延层的厚度（单位：微米）．数据如表 6.25 所示．

表 6.25 思考题 6.12 的 2^2 设计

A	B	重复					因子水平	
		I	II	III	IV		低 (−)	高 (+)
−	−	14.037	16.165	13.972	13.907	A	55%	59%
+	−	13.880	13.860	14.032	13.914			
−	+	14.821	14.757	14.843	14.878	B	短	长
+	+	14.888	14.921	14.415	14.932		(10 分钟)	(15 分钟)

(a) 估计因子效应．

(b) 进行方差分析．哪些因子重要？

(c) 写出一个回归方程，用于预测外延层厚度随本实验中砷流速率和沉积时间的变化情况．

(d) 分析残差．有哪些残差值得注意？

(e) 讨论你将如何处理在 (d) 中发现的潜在的异常值．

6.13 思考题 6.12 的继续．用思考题 6.12 的 (c) 部分中的回归模型来生成外延层厚度的响应曲面等高线图．假设得到外延层厚度为 14.5 μm 是极其重要的．你推荐什么样的砷流速率和沉积时间．

6.14 思考题 6.13 的继续．如果过程中砷流速率比沉积时间难控制，如何回答思考题 6.13 的变化？

6.15 镍钛合金用来制造喷气涡轮飞机发动机的零件．制成品上的裂缝是一个潜在的严重问题，因为它导致不可修复的报废．生产部门进行一项实验，以便确定 4 个因子对裂缝的效应．4 个因子是浇注温度 (A)、钛含量 (B)、热处理方法 (C) 以及所用的晶粒细化剂量 (D)．采用有两次重复的 2^4 设计．对试样按例行试验进行，并测量引起的裂缝长度（单位：10^{-5} 毫米）．数据如下：

(a) 估计因子效应．哪些因子效应显得较大？

(b) 进行方差分析．任一因子都对裂缝起作用吗？用 $\alpha = 0.05$.

(c) 写出一个可以预测裂缝长度的回归模型，裂缝长度是你在 (b) 中识别的显著主效应和交互作用的函数．

(d) 分析此实验残差．

(e) 有证据表明任一因子都对裂缝的变异性起作用吗？

(f) 进行生产时，你的建议是什么？用交互作用和/或主效应图帮助你得出结论．

A	B	C	D	处理组合	重复	
					I	II
−	−	−	−	(1)	7.037	6.376
+	−	−	−	a	14.707	15.219
−	+	−	−	b	11.635	12.089
+	+	−	−	ab	17.273	17.815
−	−	+	−	c	10.403	10.151
+	−	+	−	ac	4.368	4.098
−	+	+	−	bc	9.360	9.253
+	+	+	−	abc	13.440	12.923
−	−	−	+	d	8.561	8.951
+	−	−	+	ad	16.867	17.052
−	+	−	+	bd	13.876	13.658
+	+	−	+	abd	19.824	19.639
−	−	+	+	cd	11.846	12.337
+	−	+	+	acd	6.125	5.904
−	+	+	+	bcd	11.190	10.935
+	+	+	+	$abcd$	15.653	15.053

6.16 思考题 6.15 的继续. 思考题 6.15 描述的实验中的一个变量, 热处理方法 (C), 是一个类别变量. 假设其余因子是连续的.

(a) 写出两个预测裂缝长度的回归模型, 每一个都是热处理变量的一个水平. 你注意到这两个方程有哪些不同之处吗?

(b) 作出 (a) 中两个回归模型的合适的响应曲面等高线图.

(c) 如果用加热处理法 $C=+$, 你推荐 A, B, D 的哪组条件?

(d) 假设希望用加热处理法 $C=-$, 重复 (c) 部分.

6.17 一个实验者做了 2^4 设计的一次重复. 计算得到下面的效应估计:

$A=76.95$	$AB=-51.32$	$ABC=-2.82$
$B=-67.52$	$AC=11.69$	$ABD=-6.50$
$C=-7.84$	$AD=9.78$	$ACD=10.20$
$D=-18.73$	$BC=20.78$	$BCD=-7.98$
	$BD=14.74$	$ABCD=-6.25$
	$CD=1.27$	

(a) 作出这些效应的正态概率图.

(b) 基于 (a) 中效应的图像, 确定一个尝试性的模型.

*6.18 考虑例 5.3 瓶子灌装实验的一个变形. 假定所用的碳酸只有两个水平, 实验是具有两次重复的 2^3 析因设计. 数据如下:

(a) 分析实验数据. 哪些因子显著影响灌装高度的偏差?

(b) 分析实验的残差. 有迹象表明模型的不适合性吗?

(c) 根据工序变量的重要性, 得到预测灌装高度偏差的模型. 用这个模型构作等高线图, 用于帮助解释实验的结果.

(d) 在 (a) 中, 你大概注意到交互作用比较显著. 如果在模型中没有包含交互作用, 现在包含它, 再做一次分析. 这样做有什么不同? 如果在 (a) 中你选择了包含交互作用, 现在移去它重新分析. 交互作用项又产生了什么不同?

	规范因子			灌装高度偏差			因子水平	
试验	A	B	C	重复 1	重复 2		低 (−1)	高 (+1)
1	−	−	−	−3	−1	A(%)	10	12
2	+	−	−	0	1	B(psi)	25	30
3	−	+	−	−1	0	C(b/m)	200	250
4	+	+	−	2	3			
5	−	−	+	−1	0			
6	+	−	+	2	1			
7	−	+	+	1	1			
8	+	+	+	6	5			

*6.19 我一直想提高高尔夫球成绩. 由于一个高尔夫球手在他的 35% 到 45% 的击打中用到推杆; 为了提高高尔夫成绩, 改进打球入洞动作是合乎逻辑且简单的方法 ("谁能轻轻击球入洞, 谁就是胜利者."——Willie Paeks, 1864 − 1925, 两次英国公开赛的冠军). 进行一个实验来研究 4 个因子对打球入洞准确度的影响. 这些设计因子包括: 击球入洞的距离、推杆的类型、入洞方向线 (曲线还是直线)、击球入洞的倾斜面 (水平还是下坡). 响应变量是球静止后离球洞的距离. 一个高尔夫手进行有 7 次重复的 2^4 析因设计的实验, 所有击球入洞的次序都是随机的. 结果如下.

设计因子				球静止后离球洞的距离 (重量)						
击球入洞的距离	推杆的类型	入洞方向线	击球入洞的倾斜面	1	2	3	4	5	6	7
10	椎形杆头	直线	水平	10.0	18.0	14.0	12.5	19.0	16.0	18.5
30	椎形杆头	直线	水平	0.0	16.5	4.5	17.5	20.5	17.5	33.0
10	中空式铁杆	直线	水平	4.0	6.0	1.0	14.5	12.0	14.0	5.0
30	中空式铁杆	直线	水平	0.0	10.0	34.0	11.0	25.5	21.5	0.0
10	椎形杆头	曲线	水平	0.0	0.0	18.5	19.5	16.0	15.0	11.0
30	椎形杆头	曲线	水平	5.0	20.5	18.0	20.0	29.5	19.0	10.0
10	中空式铁杆	曲线	水平	6.5	18.5	7.5	6.0	0.0	10.0	0.0
30	中空式铁杆	曲线	水平	16.5	4.5	0.0	23.5	8.0	8.0	8.0
10	椎形杆头	直线	下坡	4.5	18.0	14.5	10.0	0.0	17.5	6.0
30	椎形杆头	直线	下坡	19.5	18.0	16.0	5.5	10.0	7.0	36.0
10	中空式铁杆	直线	下坡	15.0	16.0	8.5	0.0	0.5	9.0	3.0
30	中空式铁杆	直线	下坡	41.5	39.0	6.5	3.5	7.0	8.5	36.0
10	椎形杆头	曲线	下坡	8.0	4.5	6.5	10.0	13.0	41.0	14.0
30	椎形杆头	曲线	下坡	21.5	10.5	6.5	0.0	15.5	24.0	16.0
10	中空式铁杆	曲线	下坡	0.0	0.0	0.0	4.5	1.0	4.0	6.5
30	中空式铁杆	曲线	下坡	18.0	5.0	7.0	10.0	32.5	18.5	8.0

(a) 分析该实验的数据. 哪些因子显著影响推杆的性能?

(b) 分析该实验的残差. 有迹象表明模型的不适合性吗?

6.20 半导体产品制造过程是长而复杂的装配流程, 因此整个工厂的生产线都有多个工序中要使用点阵标识和点阵式二维码识读器. 无法识读的点阵图案会对工厂生产率造成负面影响, 因为必须要暂停加工过程直到手工输入部分数据后才能继续生产. 有一个 2^4 工厂实验用来设计激光刻码方式刻制点阵式二维码, 是刻在保护基材成型模具金属盖上的图案. 设计因子是, A = 激光强度 (9 W, 13 W),

$B=$ 激光脉冲频率 (4 000 Hz, 12 000 Hz), C = 点阵图形单元尺寸 (0.07 in, 0.12 in), D = 刻码速度 (10 in/sec, 20 in/sec), 响应变量是容错率 (UEC). 这是点阵式二维码中冗余信息比率的衡量参数. UEC 为 0 表示容错性最高, 即以最低的识读条件仍可以读出正确的解码点阵数据, UEC 为 1 表示容错性最低, 即需要最高的识读条件才可以正确识读. 一个 DMX 核对器用来测量 UEC. 该实验的数据如表 6.26 所示.

表 6.26 思考题 6.20 的 2^4 实验

标准次序	试验次序	激光功率	脉冲频率	点阵尺寸	写入速度	UEC
8	1	1.00	1.00	1.00	−1.00	0.8
10	2	1.00	−1.00	−1.00	1.00	0.81
12	3	1.00	1.00	−1.00	1.00	0.79
9	4	−1.00	−1.00	−1.00	−1.00	0.6
7	5	−1.00	1.00	1.00	−1.00	0.65
15	6	−1.00	1.00	1.00	1.00	0.55
2	7	1.00	−1.00	−1.00	−1.00	0.98
6	8	1.00	−1.00	1.00	−1.00	0.67
16	9	1.00	1.00	1.00	1.00	0.69
13	10	−1.00	−1.00	1.00	1.00	0.56
5	11	−1.00	−1.00	1.00	−1.00	0.63
14	12	1.00	−1.00	1.00	1.00	0.65
1	13	−1.00	−1.00	−1.00	−1.00	0.75
3	14	−1.00	1.00	−1.00	−1.00	0.72
4	15	1.00	1.00	−1.00	−1.00	0.98
11	16	−1.00	1.00	−1.00	1.00	0.63

(a) 分析该实验的数据. 哪些因子显著影响 UEC?

(b) 分析该实验的残差. 有迹象表明模型的不适合性吗?

6.21 重新考虑思考题 6.20 中的实验. 假设 4 个中心点是可用的而且在 4 个试验点的 UEC 响应分别是 0.98, 0.95, 0.93, 0.96. 在分析中加入弯曲性检验, 重新分析实验, 你能得出什么结果? 你对实验者有什么建议?

6.22 一家公司通过直接邮递进行销售. 做一个实验, 用于研究 3 个因子对一个特定产品的客户响应率的效应. 这 3 个因子是 $A=$ 邮件的类型 (第 3 类, 第 1 类), $B=$ 说明书的类型 (彩色, 黑白), $C=$ 提供的价格 (\$19.95, \$24.95). 邮寄给两组 8 000 个随机选择的客户, 其中每组的 1 000 个客户接受一个处理组合. 每组的客户视为一个重复. 响应变量是订单数. 实验数据如下表所示.

	规范因子			订单数		因子水平		
试验	A	B	C	重复 1	重复 2		低 (−1)	高 (+1)
1	−	−	−	50	54	A(类型)	3rd	1st
2	+	−	−	44	42	B(类型)	黑白	彩色
3	−	+	−	46	48	C(\$)	19.95	24.95
4	+	+	−	42	43			
5	−	−	+	49	46			
6	+	−	+	49	45			
7	−	+	+	47	48			
8	+	+	+	56	54			

(a) 分析该实验的数据. 哪些因子显著影响客户响应率?

(b) 分析该实验的残差. 有迹象表明模型的不适合性吗?

(c) 你对公司有什么建议?

6.23 考虑例 6.2 中单次重复的 2^4 设计. 假定所有的三因子和四因子交互作用可被忽略, 分析这些数据. 进行这一分析并和从例中所得的结论进行比较. 你认为高阶的交互作用可以被忽略的这一假定是否合理?

*6.24 在一半导体加工厂进行一个实验, 以增加产量. 研究 5 个因子, 每个取两个水平. 这些因子 (和水平) 是 A = 孔径设置 (小, 大), B = 曝光时间 (低于额定的 20%, 高于额定的 20%), C = 冲洗时间 (30 秒, 45 秒), D = 掩膜尺寸 (小, 大), E = 蚀刻时间 (14.5 分, 15.5 分). 进行如下的无重复的 2^5 设计.

(1)=7	d=8	e=8	de=6
a=9	ad=10	ae=12	ade=10
b=34	bd=32	be=35	bde=30
ab=55	abd=50	abe=52	$abde$=53
c=16	cd=18	ce=15	cde=15
ac=20	acd=21	ace=22	$acde$=20
bc=40	bcd=44	bce=45	$bcde$=41
abc=60	$abcd$=61	$abce$=65	$abcde$=63

(a) 画出效应估计的正态概率图. 哪些效应显得较大?

(b) 进行方差分析以确认你对 (a) 求得的结论.

(c) 写出与显著过程变量相关的产率的回归模型.

(d) 在正态概率纸上画残差图. 对此图满意吗?

(e) 画出残差与预测产量的关系图以及残差与各因子的关系图. 论述这些图形.

(f) 解释显著的交互作用.

(g) 你对过程运作条件有什么建议?

(h) 将本问题的 2^5 设计投影到重要因子的 2^k 设计中去. 画出此设计的草图并标明每一试验的平均产量和产量的极差. 此图对解释数据有帮助吗?

6.25 思考题 6.24 的继续. 假设实验者在最初实验的 32 次试验中增加 4 个中心点. 4 个中心点的产率分别是 68, 74, 76, 70.

(a) 包括一个纯二次项弯曲的检验, 重新分析实验.

(b) 讨论你下一步想做什么?

*6.26 在改革生产线时研究产量, 考虑 4 个因子: 时间 (A)、浓度 (B)、压力 (C) 以及温度 (D). 每个因子取两个水平. 采用单次重复的 2^4 设计, 所得数据如下表所示:

试验次序	实际试验次序	A	B	C	D	产率	因子水平		
								低 (−)	高 (+)
1	5	−	−	−	−	12	A(h)	2.5	3
2	9	+	−	−	−	18	B(%)	14	18
3	8	−	+	−	−	13	C(psi)	60	80
4	13	+	+	−	−	16	D(°C)	225.0	250
5	3	−	−	+	−	17			
6	7	+	−	+	−	15			
7	14	−	+	+	−	20			
8	1	+	+	+	−	15			
9	6	−	−	−	+	10			
10	11	+	−	−	+	25			
11	2	−	+	−	+	13			
12	15	+	+	−	+	24			
13	4	−	−	+	+	19			
14	16	+	−	+	+	21			
15	10	−	+	+	+	17			
16	12	+	+	+	+	23			

(a) 画出效应估计量的正态概率图, 哪些因子显得有较大的效应?

(b) 用 (a) 的正态概率图进行方差分析并导出误差项. 你的结论什么?

(c) 写出产率与重要过程变量关系的回归模型.

(d) 分析此实验的残差. 你的分析显示出什么潜在的问题吗?

(e) 此设计可压缩为有两次重复的 2^3 设计吗? 如果可以, 画出此设计的草图, 并在立方体的各点处表示出产量的平均值和残差. 解释这些结果.

*6.27 思考题 6.26 的继续. 用思考题 6.26(c) 部分中的回归模型得到产率的响应等高线图. 讨论该响应等高线图的实际价值.

6.28 美味胡桃巧克力小方饼实验. 作者是一位受过培训的工程师, 确信实践出真知. 我教授实验设计课已有多年, 而听众是各色各样的. 讲课时我经布置实际的实验给学员, 由他们来设计方案, 实施并分析. 他们看来很乐意这种实践, 而且从中常能学到好多知识. 本题利用亚利桑那州立大学 Gretchen Krueger 的实验结果.

有很多不同的方法来烘烤胡桃巧克力小方饼. 该实验的目的是来确定不同的烤盘材料、调料品牌及搅拌方式是如何影响小方饼的味道的. 因子水平是

因子	低 (−)	高 (+)
A= 烤盘材料	玻璃	铝
B= 搅拌方式	勺子	棒
C= 调料品牌	贵	便宜

响应变量是味道, 是一种主观性度量, 它可由若干个受试者品尝每一炉小方饼后所填的问卷而得 (问卷中可问及口味、外观、软硬、香味等). 一个 8 人试验小组品尝每一炉并填好问卷. 全部数据如下.

小方饼的批次	A	B	C	问卷测试结果							
				1	2	3	4	5	6	7	8
1	−	−	−	11	9	10	10	11	10	8	9
2	+	−	−	15	10	16	14	12	9	6	15
3	−	+	−	9	12	11	11	11	11	11	12
4	+	+	−	16	17	15	12	13	13	11	11
5	−	−	+	10	11	15	8	6	8	9	14
6	+	−	+	12	13	14	13	9	13	14	9
7	−	+	+	10	12	13	10	7	7	17	13
9	+	+	+	15	12	15	13	12	12	9	14

(a) 如果将此实验作为一个有 8 次重复的 2^3 设计, 分析数据, 并论述所得的结果.

(b) (a) 中的分析方法正确吗? 实际只有 8 炉, 我们真有 2^3 设计的 8 次重复吗?

(c) 分析味道值的平均数与标准差. 论述该结果. 该分析比 (a) 中的分析恰当吗? 为什么?

*6.29 做一个产生聚合体的实验. 研究 4 个因子温度 (A)、催化剂浓度 (B)、时间 (C) 和压力 (D). 观测两个响应: 分子重量和黏度. 设计表格和响应数据如表 6.27 所示.

(a) 只考虑分子重量的响应. 在正态概率图上画出效应估计. 什么效应显得很重要?

(b) 用方差分析确定 (a) 部分的结果. 它显示弯曲吗?

(c) 写出一个回归模型来预测分子重量, 把它表示成若干重要变量的函数的形式.

(d) 分析残差并评价模型的优劣.

(e) 用黏度响应重复 (a)~(d) 部分.

表 6.27 思考题 6.29 的 2^4 实验

试验次序	实际试验次序	A	B	C	D	分子重量	黏度	因子水平	低 (−)	高 (+)
1	18	−	−	−	−	2 400	1 400	A(℃)	100	120
2	9	+	−	−	−	2 410	1 500	B(%)	4	8
3	13	−	+	−	−	2 315	1 520	C(min)	200	30
4	8	+	+	−	−	2 510	1 630	D(psi)	60	75
5	3	−	−	+	−	2 615	1 380			
6	11	+	−	+	−	2 625	1 525			
7	14	−	+	+	−	2 400	1 500			
8	17	+	+	+	−	2 750	1 620			
9	6	−	−	−	+	2 400	1 400			
10	7	+	−	−	+	2 390	1 525			
11	2	−	+	−	+	2 300	1 500			
12	10	+	+	−	+	2 520	1 500			
13	4	−	−	+	+	2 625	1 420			
14	19	+	−	+	+	2 630	1 490			
15	15	−	+	+	+	2 500	1 500			
16	20	+	+	+	+	2 710	1 600			
17	1	0	0	0	0	2 515	1 500			
18	5	0	0	0	0	2 500	1 460			
19	16	0	0	0	0	2 400	1 525			
20	12	0	0	0	0	2 475	1 500			

*6.30 思考题 6.29 的继续. 用关于分子重量和黏质的回归模型回答下列的问题.

(a) 构造分子重量的一个响应等高线图. 在什么方向调整过程变量可以增加分子重量?

(b) 构造黏度的一个响应等高线图. 在什么方向调整过程变量可以减少黏度?

(c) 如果生产一种产品, 要求分子重量在 2 400~2 500 之间, 且有最低可能的黏度, 你对操作条件有什么建议?

6.31 考虑例 6.2 单次重复的 2^4 设计. 设在中心点 (0, 0, 0, 0) 处做 5 次试验并观测得下列响应: 93, 95, 91, 89, 96. 检验此实验的弯曲性. 解释这些结果.

6.32 在 2^k 析因设计中有一个缺失值. 在 2^k 设计中, 由于测量设备、发生了故障, 或者测试遭到了损坏等原因, 造成了缺失一个观测值的事情并不少见. 如果设计是 n 次 ($n > 1$) 重复的, 则可用第 5 章讨论过的一些方法. 不过, 对一个无重复的析因设计 ($n = 1$) 说来, 必须用另外的方法. 一个合乎逻辑的方法是用使高阶交互作用的对照为零的数来估计这一缺失值. 将此法应用于例 6.2 中的实验. 假定缺失了试验 ab, 将此结果与例 6.2 的结果进行比较.

6.33 工程师做一个实验来研究 4 个因子对加工零件表面粗糙度的影响. 这些因子 (以及它们的水平) 是 A= 工具角度 (12°, 15°), B= 切割液体黏度 (300, 400), C= 进料速度 (10 in/min, 15 in/min), D= 切割液体冷却机的使用 (否, 是). 实验的数据 (因子规范为一般的 −1, +1 水平) 列在下表中.

(a) 估计因子效应. 在正态概率图上画出效应估计, 选择一个试验性的模型.

(b) 拟合在 (a) 部分被识别的模型并分析残差. 有迹象表明模型的不适合性吗?

(c) 用 $1/y$ 作为响应变量, 重复 (a)、(b) 部分中的分析. 有迹象表明这个变换有用吗?

(d) 用规范变量拟合一个可以用于预测表面粗糙度的模型. 将这个方程转换成自然变量的模型.

试验	A	B	C	D	表面粗糙度	试验	A	B	C	D	表面粗糙度
1	−	−	−	−	0.003 40	9	−	−	−	+	0.003 36
2	+	−	−	−	0.003 62	10	+	−	−	+	0.003 44
3	−	+	−	−	0.003 01	11	−	+	−	+	0.003 08
4	+	+	−	−	0.001 82	12	+	+	−	+	0.001 84
5	−	−	+	−	0.002 80	13	−	−	+	+	0.002 69
6	+	−	+	−	0.002 90	14	+	−	+	+	0.002 84
7	−	+	+	−	0.002 52	15	−	+	+	+	0.002 53
8	+	+	+	−	0.001 60	16	+	+	+	+	0.001 63

*6.34 硅晶片的电阻系数受几个因子影响. 运行在关键工序步骤的 2^4 析因设计的结果列在下表:

试验	A	B	C	D	电阻系数	试验	A	B	C	D	电阻系数
1	−	−	−	−	1.92	9	−	−	−	+	1.60
2	+	−	−	−	11.28	10	+	−	−	+	11.73
3	−	+	−	−	1.09	11	−	+	−	+	1.16
4	+	+	−	−	5.75	12	+	+	−	+	4.68
5	−	−	+	−	2.13	13	−	−	+	+	2.16
6	+	−	+	−	9.53	14	+	−	+	+	9.11
7	−	+	+	−	1.03	15	−	+	+	+	1.07
8	+	+	+	−	5.35	16	+	+	+	+	5.30

(a) 估计因子效应. 在正态概率图上画出效应估计, 选择一个试验性的模型.

(b) 拟合在 (a) 部分被识别的模型. 并分析残差. 有迹象表明模型的不适合性吗?

(c) 用 $\ln(y)$ 作为响应变量重复 (a)、(b) 部分中的分析. 有迹象表明这个变换有用吗?

(d) 用规范变量拟合一个可以用于预测电阻系数的模型.

6.35 思考题 6.34 的继续. 假如: 实验者在思考题 6.34 的试验中同时进行 4 个中心点的试验. 中心点电阻系数测量值是 8.15, 7.63, 8.95 和 6.48. 合并中心点, 重新分析该实验, 你能得出什么结论?

6.36 通常在 2^k 析因设计中拟合的回归模型用于预测设计空间中感兴趣的点.

(a) 计算预测响应 $\hat{y}$ 在设计空间中的点 $x_1, x_2, \cdots, x_k$ 上的方差. **提示:** x 是规范变量, 假定 2^k 设计在每个设计点等量重复 n 次, 回归系数 $\hat{\beta}$ 的方差是 $\sigma^2/(2^k n)$, 且任意一对回归系数的协方差是零.

(b) 用 (a) 中的结果, 计算在设计空间中的点 $x_1, x_2, \cdots, x_k$ 上的关于真实平均响应的 $100(1-\alpha)\%$置信区间.

6.37 **分层模型**. 在选择模型时, 我们已多次运用分层原则, 也就是, 模型中包含不显著的低阶项, 这是因为它们被包含在显著的更高阶项中. 分层不是一个在所有情况下都必须遵循的绝对原则. 用思考题 6.1 中的模型说明, 思考题 6.1 中要求包括一个不显著的主效应以满足分层. 使用思考题 6.1 的数据,

(a) 拟合分层模型和不分层模型.

(b) 计算两个模型的 PRESS 统计量、调整的 R^2 以及均方误差.

(c) 计算立方体各角点处 $(x_1 = x_2 = x_3 = \pm 1)$ 的平均响应估计的 95%置信区间. **提示:** 使用思考题 6.36 的结果.

(d) 根据所作的分析, 你选择哪一个模型?

第 7 章　2^k 析因实验的区组设计和混区设计

本章纲要

7.1　引言
7.2　重复的 2^k 析因实验的区组设计
7.3　2^k 析因实验的混区设计
7.4　二区组的 2^k 析因实验的混区设计
7.5　区组化重要性的另一个例证
7.6　四区组的 2^k 析因实验的混区设计
7.7　2^p 个区组的 2^k 析因实验的混区设计
7.8　部分混区设计
第 7 章补充材料
S7.1　区组设计中的误差项
S7.2　区组设计的预测公式
S7.3　试验顺序是重要的

7.1　引　　言

在许多情况下, 一般不可能在同样条件下进行一个 2^k 析因实验的所有试验. 例如, 一批原材料可能不足以做所有要求的试验. 另外, 实际中遇到许多问题时, 可以通过细分实验条件使得每个处理是等效的 (即稳健的). 例如, 化学工程师可以用不同批次的原材料做一些小型试验, 因为他知道在实际的大规模生产过程中可能用到不同质量等级的不同批次的原材料.

此时, 所采用的设计技术为**区组化**. 第 4 章引入了区组化方法, 你也许会发现它有助于阅读第 4 章的引言材料. 我们在第 5 章中也讨论了区组化的一般析因设计. 本章在第 4 章概念的基础上, 侧重介绍 2^k 析因设计区组化的几种特殊方法.

7.2　重复的 2^k 析因实验的区组设计

假定 2^k 析因设计已重复 n 次. 和第 5 章中讨论的一样, 我们显示了如何在区组中进行一般的析因设计. 如果有 n 次重复, 则每一组非同质条件定义为一个区组, 在每个区组中进行一次重复. 每个区组 (即重复) 的试验按随机次序进行. 设计的分析与任一区组化的析因实验的分析类似, 参见 5.6 节中的讨论.

例 7.1　考虑在 6.2 节中首次描述的化学过程实验. 假定一批原材料只能做 4 个试验. 因此, 对该设计进行 3 次重复则需要 3 批原材料. 表 7.1 显示了该设计, 其中原材料的每一批次对应于一个区组.

表 7.1　三区组的化学过程实验

	区组1	区组2	区组3
	(1)=28	(1)=25	(1)=27
	a=36	a=32	a=32
	b=18	b=19	b=23
	ab=31	ab=30	ab=29
区组总和	B_1 = 113	B_2 = 106	B_3 = 111

区组设计的方差分析如表 7.2 所示. 所有平方和的精确计算如同标准的、非区组化的 2^k 设计. 区组平方和由区组总和计算. 令 B_1, B_2, B_3 为区组总和 (见表 7.1). 那么,

$$SS_{\text{区组}}=\sum_{i=1}^{3}\frac{B_i^2}{4}-\frac{y_{\cdots}^2}{12}=\frac{(113)^2+(106)^2+(111)^2}{4}-\frac{(330)^2}{12}=6.50$$

三区组中有两个自由度. 表 7.2 说明用区组进行设计分析的结论与 6.2 节中的结论相同, 其区组效应相对较小.

表 7.2　三区组的化学过程实验的方差分析表

方差来源	平方和	自由度	均方	F_0	P 值
区组	6.50	2	3.25		
A(浓度)	208.33	1	208.33	50.32	0.000 4
B(催化剂)	75.00	1	75.00	18.12	0.005 3
AB	8.33	1	8.33	2.01	0.206 0
误差	24.84	6	4.14		
总和	323.00	11			

7.3　2^k 析因实验的混区设计

许多问题中, 在一个区组中进行一析因设计的完全重复是不可能的. **混区设计**(confounding) 是一种设计方法, 它将一个完全的析因设计安排在多个区组之内, 区组的大小比一次重复中的处理组合的个数要小. 这一方法使得关于特定处理效应 (通常是高阶的交互作用) 的信息与来自区组的信息**不可区分**或者说**混杂**. 本章集中讨论 2^k 析因设计的混区设计. 即使是由于每一区组不包含所有的处理或处理组合而出现**不完全区组设计**, 但是对于 2^k 析因系统的特殊结构, 也将有简化的分析方法.

我们考虑有 2^p 个不完全区组的 2^k 析因设计的结构和分析方法, 其中 $p<k$. 因此, 这些设计将实施在两个区组内 ($p=1$)、4 个区组内 ($p=2$)、8 个区组内 ($p=3$), 等等.

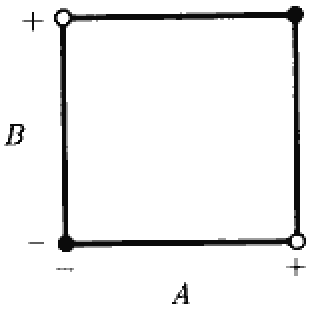
7.4　二区组的 2^k 析因实验的混区设计

假定要实施一个单次重复的 2^2 设计. 比如说, $2^2=4$ 个处理组合的每一个需用一定量的原材料, 而每批原材料只够供给两个处理组合去试验. 这样一来, 就需要两批原材料. 如果把原材料的批次看作区组, 则我们必须把这 4 种处理组合中的两种分派给每个区组.

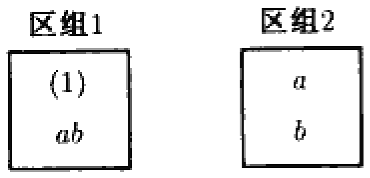
图 7.1 表示对此问题的一种可能设计. 图 7.1(a) 是几何观点, 表示把相反对角线上的处理组合分派给不同的区组. 图 7.1(b) 中, 区组 1 包含处理组合 (1) 和 ab, 区组 2 包含 a 和 b. 当然, 在同一区组内的 n 个处理组合的实施顺序是随机确定的. 先实施哪一区组的试验也是随机决定的. 假定没有划分区组而来估计 A 和 B 的主效应. 由 (6.1) 式和 (6.2) 式得

$$A=\frac{1}{2}[ab+a-b-(1)],\qquad B=\frac{1}{2}[ab+b-a-(1)]$$

因为每一估计量都有来自每个区组的一个加号的处理组合和一个减号的处理组合, 所以 A 和 B 都不受划分区组的影响. 也就是说, 区组 1 和区组 2 之间的差异抵消了.



(a) 几何观点



(b) 在二区组内安排4次试验

图 7.1 两个区组中的 2^2 设计

今考虑 AB 的交互作用

$$AB = \frac{1}{2}[ab + (1) - a - b]$$

因为两个带加号的处理组合 $[ab$ 和 $(1)]$ 都在区组 1, 两个带减号的处理组合 (a 和 b) 都在区组 2, 所以区组效应和 AB 的交互作用是一致的. 也就是说, AB 与区组混杂了.

这一点可以从 2^2 设计的加减符号表中容易看出. 此表原先在表 6.2 中给出, 为方便起见, 重列于表 7.3. 由此表可见, AB 上带加号的处理组合都安排在区组 1, 而 AB 上带减号的处理组合都安排在区组 2. 这种方式可用来把任一效应 (A, B, AB) 与区组相混. 例如, 如果 (1) 和 b 安排到区组 1, a 和 ab 安排到区组 2, 则主效应 A 与区组相混. 通常的实践是把高阶交互作用与区组相混.

表 7.3 2^2 设计的加减符号表

处理组合	析因效应				
	I	A	B	AB	区组
(1)	+	−	−	+	2
a	+	+	−	−	1
b	+	−	+	−	1
ab	+	+	+	+	2

这个方案可用来混杂任意二区组中的 2^k 设计. 作为第二个例子, 考虑二区组的 2^3 设计. 设我们想把三因子交互作用 ABC 与区组相混. 由表 7.4 的加减符号表, 我们把 ABC 中带减号的处理组合分在区组 1, ABC 中带加号的处理组合分在区组 2. 所得的设计如图 7.2 所示. 我们再次强调, 在一区组内的处理组合的试验顺序是随机的.

表 7.4 2^3 设计的加减符号表

处理组合	析因效应								
	I	A	B	AB	C	AC	BC	ABC	区组
(1)	+	−	−	+	−	+	+	−	1
a	+	+	−	−	−	−	+	+	2
b	+	−	+	−	−	+	−	+	2
ab	+	+	+	+	−	−	−	−	1
c	+	−	−	+	+	−	−	+	2
ac	+	+	−	−	+	+	−	−	1
bc	+	−	+	−	+	−	+	−	1
abc	+	+	+	+	+	+	+	+	2

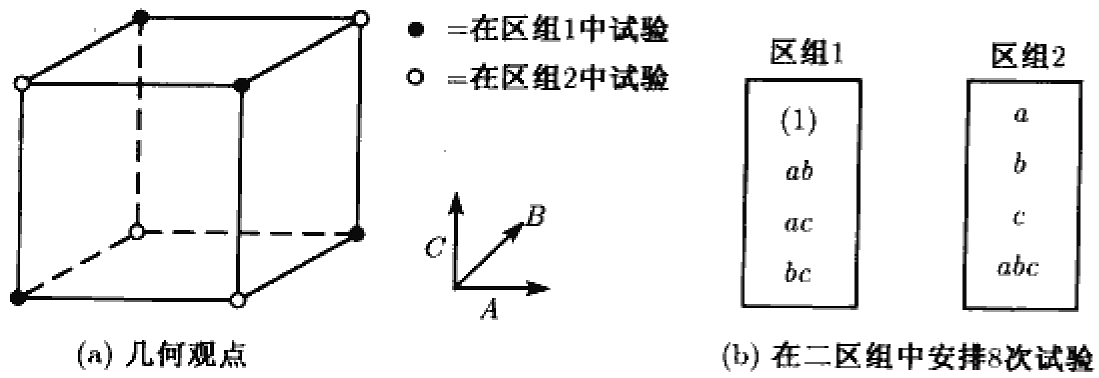


图 7.2 ABC 与二区组混杂的 2^3 设计

1. 构造区组的其他方法

有另外的方法构造这类设计. 利用线性组合

$$L = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k \tag{7.1}$$

其中 x_i 是一特定处理组合中第 i 个因子的水平, α_i 是第 i 个因子在被相混的效应中的分量. 对 2^k 系统说来, 有 $\alpha_i = 0$ 或 1, $x_i = 0$(低水平) 或 $x_i = 1$(高水平). (7.1) 式称为**定义对照**. 产生相同 L 值 (mod 2) 的处理组合放在同一区组内. 因为 L 的值 (mod 2) 只可能是 0 和 1, 这就把 2^k 个处理组合准确地分在两个区组中.

为了说明这一方法, 考虑 ABC 与区组相混的 2^3 设计. 此时, x_1 对应于 A, x_2 对应 B, x_3 对应于 C, 以及 $\alpha_1 = \alpha_2 = \alpha_3 = 1$. 于是, 对应于 ABC 的定义对照是

$$L = x_1 + x_2 + x_3$$

处理组合 (1) 在 (0, 1) 记号法中写为 000, 因此

$$L = 1(0) + 1(0) + 1(0) = 0 = 0(\text{mod } 2)$$

类似地, 处理组合 a 是 100, 得

$$L = 1(1) + 1(0) + 1(0) = 1 = 1(\text{mod } 2)$$

于是 (1) 和 a 将在不同的区组中进行试验. 对其余的处理组合, 我们有

$$\begin{aligned}
b: L &= 1(0) + 1(1) + 1(0) = 1 = 1(\text{mod}\,2)\\
ab: L &= 1(1) + 1(1) + 1(0) = 2 = 0(\text{mod}\,2)\\
c: L &= 1(0) + 1(0) + 1(1) = 1 = 1(\text{mod}\,2)\\
ac: L &= 1(1) + 1(0) + 1(1) = 2 = 0(\text{mod}\,2)\\
bc: L &= 1(0) + 1(1) + 1(1) = 2 = 0(\text{mod}\,2)\\
abc: L &= 1(1) + 1(1) + 1(1) = 3 = 1(\text{mod}\,2)
\end{aligned}$$

这样, (1), ab, ac, bc 分在区组 1, 而 a, b, c, abc 分在区组 2. 这与由加减符号表所得的图 7.2 所示的设计相同.

也可以使用另一种方法构造这些设计. 包含处理组合 (1) 的区组叫做**主区组**(principal block). 在此区组内的处理组合具有一个有用的群论性质, 即它们关于乘法模 2 构成一群. 这就是说, 主

区组内的任一元素 [除 (1) 以外] 都可由主区组内另外两个元素相乘模 2 而得到. 例如, 考虑 ABC 被混杂的 2^3 设计的主区组, 如图 7.2 所示. 有

$$\begin{aligned} ab\cdot ac &= a^2bc = bc \\ ab\cdot bc &= ab^2c = ac \\ ac\cdot bc &= abc^2 = ab \end{aligned}$$

另一区组 (或多个区组) 的处理组合可以用新区组中的一个元素和主区组中的各个元素相乘模 2 而得到. 对 ABC 被混杂的 2^3 设计, 因为主区组是 (1), ab, ac, bc, 所以可知 b 在另一区组中. 这样一来, 第二个区组的元素是

$$\begin{aligned} b\cdot(1) &= b \\ b\cdot ab = ab^2 &= a \\ b\cdot ac &= abc \\ b\cdot bc = b^2c &= c \end{aligned}$$

这与前面所得的结果一致.

2. 误差的估计

当变量的个数较小时, 比方说 $k=2$ 或 3, 通常必须进行重复以求得误差的估计量. 例如, 假定需要进行有两个区组的 ABC 被混杂的 2^3 析因实验, 实验者决定进行 4 次重复. 设计如图 7.3 所示. 在每次重复中 ABC 都被混杂.

重复I		重复II		重复III		重复IV	
区组1	区组2	区组1	区组2	区组1	区组2	区组1	区组2
(1)	abc	(1)	abc	(1)	abc	(1)	abc
ac	a	ac	a	ac	a	ac	a
ab	b	ab	b	ab	b	ab	b
bc	c	bc	c	bc	c	bc	c

图 7.3 ABC 被混杂的有 4 次重复的 2^3 设计

此设计的方差分析见表 7.5, 有 32 个观测值和 31 个总的自由度. 又因为有 8 个区组, 所以和这些区组有关的必须是 7 个自由度. 7 个自由度的一种分解如表 7.5 所示. 误差平方和实际上由重复和每个效应 (A, B, C, AB, AC, BC) 之间的二因子交互作用所组成. 将交互作用看作为零, 并把它们的均方值看作误差的估计量通常是不会有问题的. 主效应和二因子交互作用都是相对均方误差进行检验的. Cochran and Cox(1957) 指出, 区组或 ABC 均方可以和由 ABC 提供的均方误差进行比较, 它实际上是由重复 × 区组提供的. 这一检验法通常是很不灵敏的.

如果资源充足, 允许混区设计做重复, 则一般说来, 在每次重复时采用稍有不同的设计区组方法会较好一些. 这一处理方法是在各次重复中混杂不同的效应, 以便求得关于所有效应的一些信息. 这样的方法称为**部分混区设计**, 7.7 节会给予讨论. 如果 k 适当大, 比方说 $k\geqslant 4$, 则经常只做一次重复就行了. 实验者通常假定高阶交互作用可被忽略并将它们的平方和组合起来作为误差. 因子效应的正态概率图对这一点十分有帮助.

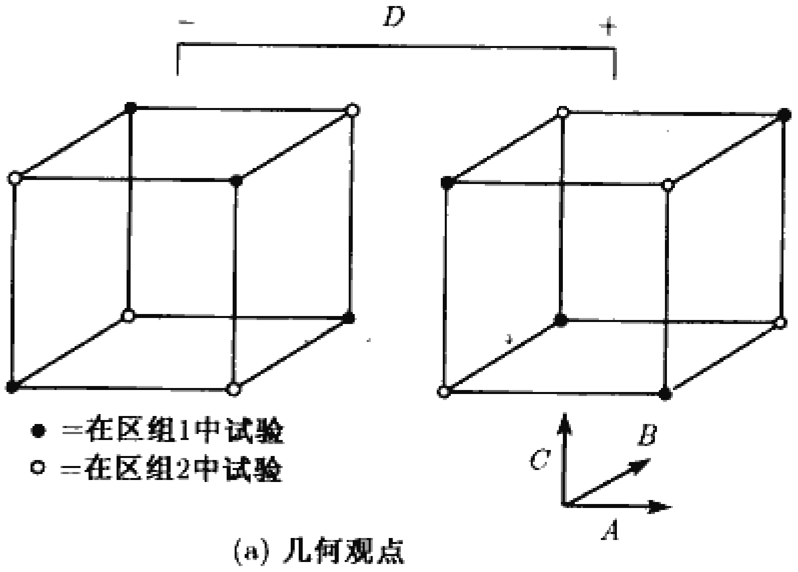
表 7.5 ABC 被混杂的有 4 次重复的 2^3 设计的方差分析表

方差来源	自由度
重复	3
区组 (ABC)	1
ABC 的误差 (重复 × 区组)	3
A	1
B	1
C	1
AB	1
AC	1
BC	1
误差 (重复 × 效应)	18
总和	31

例 7.2 考虑例 6.2 描述的情况. 回想 4 个因子: 温度 (A)、压强 (B)、甲醛的浓度 (C) 和搅拌速度 (D). 在试验性工厂中研究这些因子, 以确定影响产品渗透率的效应. 我们用这个实验来说明在一个无重复设计中区组化和混杂的思想. 我们对最初的实验作了两处修改. 首先, 假定一批原材料不能进行全部 $2^4 = 16$ 个处理组合的试验. 一批原材料只能进行 8 个处理组合的试验, 所以, 采用二区组的 2^4 混区设计看来是恰当的. 将高阶交互作用 $ABCD$ 与区组混杂也是很自然的. 定义对照是

$$L = x_1 + x_2 + x_3 + x_4$$

容易证明, 此设计如图 7.4 所示. 另外, 可以检查表 6.12. 观测被安排到区组 1 的 $ABCD$ 列中的标为 "+" 的处理组合, 以及区组 2 的 $ABCD$ 列中的标为 "−" 的处理组合.



(a) 几何观点

区组1	区组2
(1)=25	a=71
ab=45	b=48
ac=40	c=68
bc=60	d=43
ad=80	abc=65
bd=25	bcd=70
cd=55	acd=86
$abcd$=76	abd=104

(b) 在二区组中安排16次试验

图 7.4 例 7.2 的二区组的 2^4 设计

我们作的第二处修改是引入**区组效应**, 以说明区组的作用. 假定我们选择两批原材料来进行实验, 其中之一质量较差, 因此, 用这批材料得到的平均响应低于那些用好的质量的材料 20 个单位. 质量差的批次为区组 1, 质量好的批次为区组 2(哪个批次为区组 1, 哪个批次为区组 2 无关紧要). 先做区组 1 的试验 (当然, 区组里的 8 个试验是按随机次序进行的), 但是, 响应低于那些用好的质量材料的 20 个单位. 图 7.4b 显示了响应的结果 —— 注意到它们是可以由例 6.2 给出的原始观测减去区组效应得到的. 即, 处理组合 (1) 的原始响应为 45, 图 7.4b 中报告为 (1)=25(=45−20). 该区组的其他响应也可以类似得到. 做完区组 1 的试验后, 接着做区组 2 的 8 个试验. 这个批次的原材料没有问题, 所以, 响应和它们最初在例 6.2 中的非常一致.

表 7.6 所列的是例 6.2 的 "修改" 观点的效应估计. 注意到 4 个主效应、6 个二因子交互作用、4 个三因子交互作用的估计和没有区组效应的例 6.2 得到的效应估计一样. 计算效应估计的通常比例后, 可以看出因子 A, C, D 和 AC 交互作用以及 AD 交互作用是重要效应, 与原始实验一样. (读者应当确认这一点.)

表 7.6 例 7.2 中的区组化 2^4 设计的效应估计

模型项	回归系数	效应估计	平方和	百分比贡献率
A	10.81	21.625	1 870.562 5	26.30
B	1.56	3.125	39.062 5	0.55
C	4.94	9.875	390.062 5	5.49
D	7.31	14.625	855.562 5	12.03
AB	0.062	0.125	0.062 5	< 0.01
AC	−9.06	−18.125	1 314.062 5	18.48
AD	8.31	16.625	1 105.562 5	15.55
BC	1.19	2.375	22.562 5	0.32
BD	−0.19	−0.375	0.562 5	< 0.01
CD	−0.56	−1.125	5.062 5	0.07
ABC	0.94	1.875	14.062 5	0.20
ABD	2.06	4.125	68.062 5	0.96
ACD	−0.81	−1.625	10.562 5	0.15
BCD	−1.31	−2.625	27.562 5	0.39
区组 ($ABCD$)		−18.625	1 387.562 5	19.51

$ABCD$ 的交互效应是怎样的呢? 原始实验 (例 6.2) 中该效应估计为 $ABCD= 1.375$. 而在目前的例子中, $ABCD$ 交互效应的估计为 $ABCD = -18.625$. 因为 $ABCD$ 与区组混杂了, 用**原始交互效应**(1.375) 加上**区组效应**(−20) 估计 $ABCD$ 交互作用, 所以 $ABCD = 1.375 + (-20) = -18.625$. (你是否明白, 为什么区组效应为 −20?) 区组效应可以用两区组平均响应间的差直接计算, 即

$$\text{区组效应} = \bar{y}_{\text{区组 }1} - \bar{y}_{\text{区组 }2} = \frac{406}{8} - \frac{555}{8} = \frac{-149}{8} = -18.625$$

当然, 这个效应确实估计了区组 $+ABCD$.

表 7.7 描述了该实验的方差分析表. 模型中包含了大的估计的效应, 区组平方和为

$$SS_{\text{区组}} = \frac{(406)^2 + (555)^2}{8} - \frac{(961)^2}{16} = 1\ 387.562\ 5$$

该实验的结论十分符合没有区组效应的例 6.2 的结论. 如果没有在区组中进行实验, 且如果大小为 -20 的效应影响了前面的 8 个试验 (它们可能按照随机方式进行的, 因为 16 个试验在非区组设计中是按照随机次序进行的), 结果可能就不同.

表 7.7 例 7.2 的方差分析表

方差来源	平方和	自由度	均 方	F_0	P 值
区组 ($ABCD$)	1 387.562 5	1			
A	1 870.562 5	1	1 870.562 5	89.76	$< 0.000\ 1$
C	390.062 5	1	390.062 5	18.27	0.001 9
D	855.562 5	1	855.562 5	41.05	0.000 1
AC	1 314.062 5	1	1 314.062 5	63.05	$< 0.000\ 1$
AD	1 105.562 5	1	1 105.562 5	53.05	$< 0.000\ 1$
误差	187.562 5	9	20.840 3		
总和	7 111.437 5	15			

7.5 区组化重要性的另一个例证

区组化是一项非常有用且非常重要的设计技术. 第 4 章已经指出区组化能够减少实验中的噪声, 在这种实验中, 实验者通常考虑该实验的讨厌因子 (有时可能是区组) 的影响.

为了说明实验者在必须区组化而没有区组化时可能发生的情形, 考虑 7.4 节中例 7.2 的一个变形. 该例利用例 6.2 中的 2^4 无重复析因实验. 构造的设计为在两个区组中各做 8 次试验,

插入一个大小为 −20 的"区组效应"即讨厌因子效应, 它影响区组 1 的所有观测值 (见图 7.4). 假定我们对设计没有区组, 从而 −20 的讨厌因子效应影响了前 8 个观测值 (随机的或按次序进行的). 修改的数据列在表 7.8 中.

表 7.8　例 7.2 的修改数据

试验顺序	标准顺序	因子 A: 温度	因子 B: 压强	因子 C: 浓度	因子 D: 搅拌速率	响应: 渗透率
8	1	−1	−1	−1	−1	25
11	2	1	−1	−1	−1	71
1	3	−1	1	−1	−1	28
3	4	1	1	−1	−1	45
9	5	−1	−1	1	−1	68
12	6	1	−1	1	−1	60
2	7	−1	1	1	−1	60
13	8	1	1	1	−1	65
7	9	−1	−1	−1	1	23
6	10	1	−1	−1	1	80
16	11	−1	1	−1	1	45
5	12	1	1	−1	1	84
14	13	−1	−1	1	1	75
15	14	1	−1	1	1	86
10	15	−1	1	1	1	70
4	16	1	1	1	1	76

图 7.5 是修改过的实验中因子效应的正态概率图. 尽管该图表面上与第 6 章给出的实验的最初分析没有太多的不一致 (见图 6.11), 但重要的交互作用之一, AD, 并没有识别出来. 结果, 我们不能发现作为解决最初问题的关键之一的这个重要效应. 回顾第 4 章, 我们称区组化是**噪声缩减技术**. 如果不能区组化, 由讨厌变量效应增加的变异就分布在其他设计因子中.

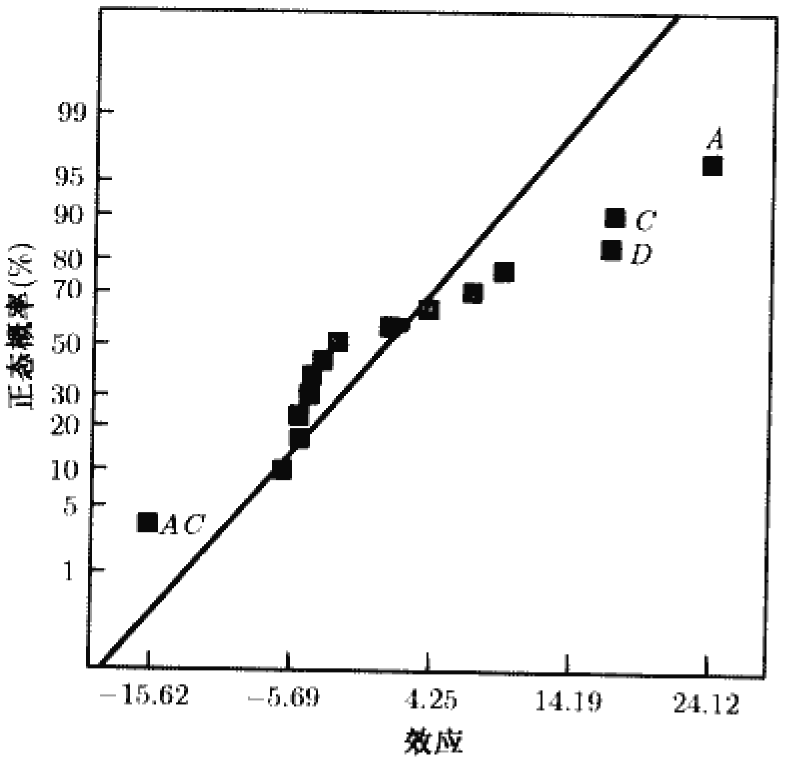


图 7.5　表 7.8 的数据的正态概率图

误差估计中也包含讨厌的变异性的一部分. 根据表 7.8 的数据, 模型的残差均方大约为 109, 它比基于原始数据的残差均方大了好几倍 (见表 6.13).

7.6 四区组的 2^k 析因实验的混区设计

可以构造与 4 个区组相混的 2^k 析因混区设计, 每个区组有 2^{k-2} 个观测值. 当因子的个数适当大 (比方说 $k \geqslant 4$), 以及区组相对地小时, 此类设计特别有用.

作为例子, 我们考虑 2^5 设计. 如果每一区组只能容纳 8 个试验, 则须用 4 个区组. 这一设计的构造相对说来是直截了当的. 选取两个与区组相混的效应, 比方说 ADE 和 BCE, 相应的两个定义对照分别是

$$L_1 = x_1 + x_4 + x_5, \qquad L_2 = x_2 + x_3 + x_5$$

于是, 每一处理组合将得出 $L_1 (\text{mod } 2)$ 和 $L_2 (\text{mod } 2)$ 的一对特定值, 也就是, $(L_1, L_2)=(0, 0)$, 或 $(0, 1)$, 或 $(1, 0)$, 或 $(1, 1)$. (L_1, L_2) 取值相同的处理组合分派在同一个区组中. 在我们的例子中,

对于 $(1), ad, bc, abcd, abe, ace, cde, bde$, 有 $L_1 = 0, L_2 = 0$

对于 $a, d, abc, bcd, be, abde, ce, acde$, 有 $L_1=1$, $L_2=0$

对于 $b, abd, c, acd, ae, de, abce, bcde$, 有 $L_1=0$, $L_2=1$

对于 $e, ade, bce, abcde, ab, bd, ac, cd$, 有 $L_1=1$, $L_2=1$

这些处理组合分派在 4 个不同的区组中. 完整的设计见图 7.6.

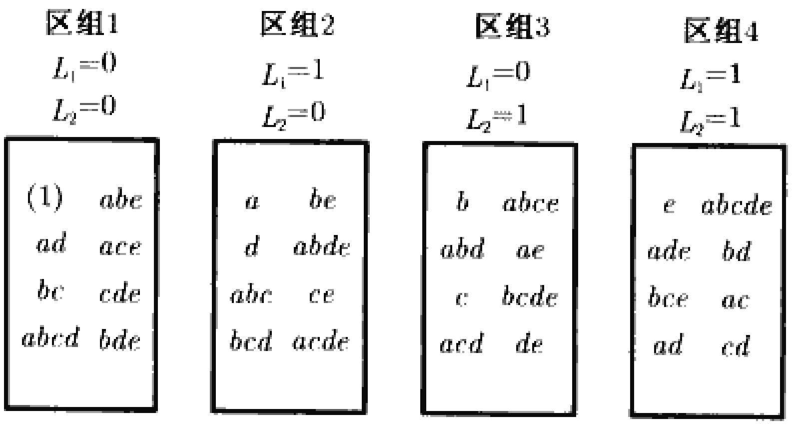


图 7.6 ADE、BCE 以及 $ABCD$ 混杂的四区组中的 2^5 设计

稍为思考一下, 就会认识到除了 ADE 和 BCE 之外, 还必须将另一效应与区组相混. 因为有 4 个区组, 其自由度为 3, 而 ADE 和 BCE 各只有 1 个自由度, 显然, 必须再要混杂一个自由度为 1 的效应. 这一效应是 ADE 和 BCE 的**广义交互作用**(generalized interaction), 定义为 ADE 与 BCE 模 2 的乘积. 这样, 在我们的例子中, 广义交互作用 $(ADE)(BCE) = ABCDE^2 = ABCD$ 亦与区组相混. 利用 2^5 设计的加减符号表 [如, 在 Davies(1956) 的书中] 容易证实这一点. 与这类表对照一下就会发现, 分派给各个区组的处理组合如下:

处理组合在	ADE 的符号	BCE 的符号	$ABCD$ 的符号
区组 1	−	−	+
区组 2	+	−	−
区组 3	−	+	−
区组 4	+	+	+

注意, 一个特定区组内的任意两个效应 (例如, ADE 和 BCE) 的符号的乘积就得同一区组内另一效应 (此时是 $ABCD$) 的符号. 这样一来, ADE、BCE 以及 $ABCD$ 都与区组相混杂.

7.4 节中所述的主区组的群论性质仍然成立. 例如, 主区组内两个处理组合的乘积就得出主区组的另一个元素. 也就是,

$$ad \cdot bc = abcd, \qquad abe \cdot bde = ab^2de^2 = ad$$

如此类推, 如要构造另一区组, 选取不在主区组内的一个处理组合 (例如, b), 并将 b 乘以主区组内各个处理组合即可. 这就得出

$$b \cdot (1) = b, \quad b \cdot ad = abd, \quad b \cdot bc = b^2c = c, \quad b \cdot abcd = ab^2cd = acd$$

如此类推, 这就得出区组 3 内的 8 个处理组合. 实际上, 主区组也可以从定义对照和群论性质来求得, 而其余的区组用上述方法所示的处理组合确定.

构造有 4 个区组的 2^k 混区设计的一般方法是, 选取生成区组的两个效应, 自动地得到被混杂的第 3 个效应, 它是前两个效应的广义交互作用. 这样一来, 用两个定义对照 (L_1, L_2) 和主区组的群论性质来构造这一设计. 在选取与区组相混的效应时, 要小心谨慎, 不要把所感兴趣的效应混进去. 例如, 在 2^5 设计中, 可能选取 $ABCDE$ 和 ABD 与区组相混, 这就自动地混进了 CE, 它可能也是一个感兴趣的效应. 一个较好的选择是选取 ADE 和 BCE 与区组相混, 它自动地混进 $ABCD$. 牺牲三因子交互作用 ADE 和 BCE 的信息来替代二因子交互作用 CE 更为可取.

7.7　2^p 个区组的 2^k 析因实验的混区设计

推广上述方法以构造 $2^p(p < k)$ 个区组中的 2^k 析因混区设计, 其中每一区组包含 2^{k-p} 个试验. 我们选取 p 个独立的效应与区组相混, 此处所谓 "独立的", 意即所选取的效应不是其他效应的广义交互作用. 区组可以用与这 p 个效应相联系的 p 个定义对照 $L_1, L_2, \cdots, L_p$ 来生成. 另外, 还有 $2^p - p - 1$ 个其他的效应与区组相混, 这些效应是原先选取的 p 个独立效应的广义交互作用. 要小心选取与区组相混的效应, 不要牺牲了可能有意义的效应的信息.

这些设计的统计分析是较为简单的. 所有效应平方和的计算和没有区组一样. 区组的平方和是将所有的区组相混的效应的平方和加起来.

显然, 选择用来生成区组的 p 个效应是关键性的, 因为混区设计的构造直接依赖于它们. 表 7.9 列出了有用的设计. 为说明此表的用法, 假定我们要构造 $2^3 = 8$ 个区组中的 2^6 混区设计, 每个区组有 $2^3 = 8$ 个试验. 表 7.9 指出, 我们可选取 $ABEF$、$ABCD$ 和 ACE 作为 $p=3$ 个独立的效应来生成区组. 其余的 $2^p - p - 1 = 2^3 - 3 - 1 = 4$ 个与区组相混的效应是这 3 个效应的广义交互作用; 也就是,

$$\begin{aligned}(ABEF)(ABCD) &= A^2B^2CDEF = CDEF\\(ABEF)(ACE) &= A^2BCE^2F = BCF\\(ABCD)(ACE) &= A^2BC^2ED = BDE\\(ABEF)(ABCD)(ACE) &= A^3B^2C^2DE^2F = ADF\end{aligned}$$

在思考题 7.11 中要求读者生成此设计的 8 个区组.

表 7.9 2^k 析因设计的区组安排建议表

因子数, k	区组数, 2^p	区组大小, 2^{k-p}	选择效应生成区组	与区组混杂的交互作用
3	2	4	ABC	ABC
	4	2	AB, AC	AB, AC, BC
4	2	8	$ABCD$	$ABCD$
	4	4	ABC, ACD	ABC, ACD, BD
	8	2	AB, BC, CD	$AB, BC, CD, AC, BD, AD, ABCD$
5	2	16	$ABCDE$	$ABCDE$
	4	8	ABC, CDE	$ABC, CDE, ABDE$
	8	4	ABE, BCE, CDE	$ABE, BCE, CDE, AC, ABCD, BD, ADE$
	16	2	AB, AC, CD, DE	所有二因子和四因子交互作用 (15 个效应) $ABCDEF$
6	2	32	$ABCDEF$	$ABCDEF$
	4	16	$ABCF, CDEF$	$ABCF, CDEF, ABDE$
	8	8	$ABEF, ABCD, ACE$	$ABEF, ABCD, ACE, BCF, BDE, CDEF, ADF$
	16	4	ABF, ACF, BDF, DEF	$ABF, ACF, BDF, DEF, BC, ABCD, ABDE, AD, ACDE, CE, CDF, BCDEF, ABCEF, AEF, BE$
	32	2	AB, BC, CD, DE, EF	所有二因子、四因子、六因子交互作用 (31 个效应)
7	2	64	$ABCDEFG$	$ABCDEFG$
	4	32	$ABCFG, CDEFG$	$ABCFG, CDEFG, ABDE$
	8	16	ABC, DEF, AFG	$ABC, DEF, AFG, ABCDEF, BCFG, ADEG, BCDEG$
	16	8	$ABCD, EFG, CDE, ADE$	$ABCD, EFG, CDE, ADG, ABCDEFG, ABE, BCG, CDFG, ADEF, ACEG, ABFG, BCEF, BDEG, ACF, BDF$
	32	4	ABG, BCG, CDG, DEG, EFG	$ABG, BCG, CDG, DEG, EFG, AC, BD, CE, DF, AE, BF, ABCD, ABDE, ABEF, BCDE, BCEF, CDEF, ABCDEFG, ADG, ACDEG, ACEFG, ABDFG, ABCEG, BEG, BDEFG, CFG, ADEF, ACDF, ABCF, AFG, BCDFG$
	64	2	AB, BC, CD, DE, EF, FG	所有二因子、四因子、六因子交互作用 (63 个效应)

7.8　部分混区设计

7.4 节中指出, 除非实验者有一先验的误差估计量或者愿意假定某些交互作用可以被忽略, 否则, 他们必须重复这个设计以便得到误差的估计量. 图 7.3 表示二区组中的, ABC 被混杂的 2^3 析因混区设计, 有 4 次重复. 此设计的方差分析如表 7.5 所示, 可以看出 ABC 交互作用的信息不能找出, 因为 ABC 在每次重复中都与区组相混. 这种设计称为**完全混区设计**.

现考虑如图 7.7 所示的设计. 还是有 4 次重复的 2^3 设计, 但每次重复有一个不同的交互作用与区组相混. 具体来说, 在重复 I 中 ABC 被混杂, 在重复 II 中 AB 被混杂, 在重复 III 中 BC 被混杂, 在重复 IV 中 AC 被混杂. 因此, 关于 ABC 的信息可以从重复 II、III 和 IV 的数据中获得, 关于 AB 的信息可以从重复 I、III 和 IV 获得, 关于 AC 的信息可以从重复 I、II 和 III 获得, 关于 BC 的信息可从重复 I、II 和 IV 中获得. 可以说, 我们获得关于这些交互作用的 3/4 的信息, 因为它们在 3 次重复中没有与区组相混. Yates(1937) 称比值 3/4 为**混杂效应的相对信息**. 我们称这种设计为**部分混区设计**.

重复I 混杂ABC		重复II 混杂AB		重复III 混杂BC		重复IV 混杂AC	
(1)	a	(1)	a	(1)	b	(1)	a
ab	b	c	b	a	c	b	c
ac	c	ab	ac	bc	ab	ac	ab
bc	abc	abc	bc	abc	ac	abc	bc

图 7.7　2^3 设计中的部分混杂

这一设计的方差分析如表 7.10 所示. 计算某交互作用的平方和时, 只用没有混进该交互作用的那些重复的数据. 误差平方和的组成为重复 × 主效应的平方和加上重复 × 该重复中未被混杂的交互作用平方和 (例如, 重复 × 重复 II、III 和 IV 的 ABC). 而且, 8 个区组有 7 个自由度. 通常将其分解为重复的 3 个自由度和在各次重复内的对于区组的 4 个自由度. 区组平方和的组成见表 7.10, 并可以直接由每次重复所选取的混杂的效应求得.

例 7.3　2^3 部分混区设计

考虑例 6.1, 进行一个实验以改进晶片蚀刻工序. 有 3 个因子: A = 间隙, B = 气流, C =RF 功率, 响应变量为蚀刻率. 设每个班次仅能试验 4 种处理组合. 因为蚀刻工具的性能在不同班次之间有差异, 实验者决定将班次作为区组. 这样一来, 2^3 设计的每次重复就必须分成两个区组来进行. 进行两次重复, 在重复 I 中混杂 ABC, 在重复 II 中混杂 AB. 数据如下:

重复I 混杂ABC		重复II 混杂AB	
(1)= 550	a= 669	(1)= 604	a= 650
ab= 642	b= 633	c=1 052	b= 601
ac= 749	c=1 037	ab= 635	ac= 868
bc=1 075	abc= 729	abc= 860	bc=1 063

A, B, C, AC 以及 BC 的平方和可用通常方法计算, 使用全部的 16 个观测值. 我们必须仅用重复 II 的数据来求 SS_{ABC}, 仅用重复 I 的数据求 SS_{AB}, 如下:

$$SS_{ABC}=\frac{[a+b+c+abc-ab-ac-bc-(1)]^2}{2^k n}$$
$$=\frac{[650+601+1\ 052+860-635-868-1\ 063-604]^2}{(1)(8)}=6.125\ 0$$
$$SS_{AB}=\frac{[(1)+abc-ac+c-a-b+ab-bc]^2}{2^k n}$$
$$=\frac{[550+729-749+1\ 037-669-633+642-1\ 075]^2}{(1)(8)}=3\ 528.0$$

表 7.10 部分混区 2^3 设计的方差分析表

方差来源	自由度
重复	3
重复内的区组 [即 ABC(记为 I)+AB(记为 II)+BC(记为III)+AC(记为IV)]	4
A	1
B	1
C	1
AB(由重复 I, III, IV)	1
AC(由重复 I, II, III)	1
BC(由重复 I, II, IV)	1
ABC(由重复 II, III, IV)	1
误差	17
总和	31

重复的平方和, 一般是

$$SS_{\text{重复}}=\sum_{h=1}^{n}\frac{R_h^2}{2^k}-\frac{y_{\cdots}^2}{N}=\frac{(6\ 084)^2+(6\ 333)^2}{8}-\frac{(12\ 417)^2}{16}=3\ 875.062\ 5$$

其中 R_h 是第 h 次重复中观测值的总和. 区组平方和是重复 I 中的 SS_{ABC} 与重复 II 中的 SS_{AB} 之和, 即 $SS_{\text{区组}}=458.125\ 0$.

方差分析概括在表 7.11 中. 主效应 A 和 C 以及交互作用 AC 都是重要的.

表 7.11 例 7.3 的方差分析表

方差来源	平方和	自由度	均方	F_0	P 值
重复	3 875.062 5	1	3 875.062 5	—	
重复内的区组	458.125 0	2	229.062 5	—	
A	41 310.562 5	1	41 310.562 5	16.20	0.01
B	217.562 5	1	217.562 5	0.08	0.78
C	374 850.562 5	1	374 850.562 5	146.97	< 0.001
AB(仅用重复 I)	3 528.000 0	1	3 528.000 0	1.38	0.29
AC	94 404.562 5	1	94 404.562 5	37.01	< 0.001
BC	18.062 5	1	18.062 5	0.007	0.94
ABC(仅用重复 II)	6.125 0	1	6.125 0	0.002	0.96
误差	12 752.312 5	5	2 550.462 5		
总和	531.420.937 5	15			

7.9 思 考 题

*7.1 考虑思考题 6.1 中描述的实验. 分析这个实验, 假定每次重复代表一个生产班次的一个区组.

7.2 考虑思考题 6.5 中描述的实验. 分析这个实验. 假设 4 次重复中的每一次都代表一个区组.

7.3 考虑思考题 6.15 描述的合金裂缝试验. 假定一天仅作 16 试验, 每次重复作为一次区组. 分析这个实验并得出结论.

*7.4 考虑思考题 6.1 第一次重复中的数据. 设这些观测值不能用相同的条材做试验而得. 对这些观测值建立有两个区组的混区设计, 每个区组有 4 个观测值, ABC 与区组相混. 分析这些数据.

7.5 考虑思考题 6.7 第一次重复中的数据. 构造一个混杂 $ABCD$ 的二区组的混区设计, 每个区组有 8 个观测值. 分析这些数据.

7.6 思考题 7.5 中改为假定需用 4 个区组. ABD 和 ABC(从而 CD) 与区组混杂.

*7.7 用思考题 6.24 中 2^5 设计的数据. 构造并分析二区组中的混区设计, 其中, $ABCDE$ 与区组混杂.

*7.8 思考题 7.7 中改为需用 4 个区组. 提出一个合理的混区方案.

*7.9 考虑思考题 6.24 中的 2^5 设计的数据. 假定需用混杂 $ACDE$ 和 BCD(从而 ABE) 的 4 区组的混区设计. 分析此设计所得的数据.

*7.10 考虑思考题 6.18 的灌装高度偏差实验. 假定一天只能进行一次重复. 假设工作日为区组, 分析数据.

7.11 考虑思考题 6.18 的灌装高度偏差实验. 假定每个班次只能做 4 次试验. 建立一个设计, 使得 ABC 混杂在重复 1 中, AC 混杂在重复 2 中. 分析数据并论述你的发现.

*7.12 考虑思考题 6.19 的高尔夫实验. 考虑把每次重复作为区组, 分析数据.

7.13 考虑思考题 6.20 的 2^4 设计数据, 构建并分析 $ABCD$ 与区组混杂的二区组的设计.

7.14 考虑思考题 6.22 的直接邮递试验, 假定每组顾客来自城镇的不同地方. 建议该实验的一个合理分析.

7.15 设计一个四区组中的 2^6 混区设计, 提议一个与表 7.8 不同的混区方案.

7.16 考虑一个八区组中的 2^6 混区设计, 每个区组有 8 次试验, 选取 $ABCD, ACE, ABEF$ 作为与区组相混的独立的效应. 生成这一设计. 找出其他与区组相混的效应.

*7.17 考虑混杂 AB 的二区组 2^2 设计. 用代数方法证明 $SS_{AB} = SS_{区组}$.

7.18 考虑例 7.2 的数据. 设区组 2 内所有的观测值都增加 20. 分析由此而生的数据. 估计区组效应. 对它的大小你能解释吗? 区组是一重要因子吗? 由于你对数据作了这样的变动, 有其他效应估计量受到影响吗?

7.19 设在思考题 6.1 中, 重复 I 混杂 ABC, 重复 II 混杂 AB, 重复 III 混杂 BC. 计算因子效应估计, 写出方差分析表.

7.20 假定在每次重复中 ABC 与区组相混, 重复分析思考题 6.1.

7.21 设在思考题 6.7 中, 重复 I 中混杂 $ABCD$, 重复 II 中混杂 ABC. 对此设计进行统计分析.

7.22 构造一个 2^3 混区设计. 前两次重复中混杂 ABC, 第 3 次重复中混杂 BC. 进行方差分析并论述所得的信息.

第 8 章　二水平分式析因设计

本章纲要

8.1　引言

8.2　2^k 析因设计的 $\frac{1}{2}$ 分式设计

8.2.1　定义与基本原理

8.2.2　设计分辨度

8.2.3　$\frac{1}{2}$ 分式设计的构造与分析

8.3　2^k 析因设计的 $\frac{1}{4}$ 分式设计

8.4　一般的 2^{k-p} 分式析因设计

8.4.1　设计的选择

8.4.2　2^{k-p} 分式析因设计的分析

8.4.3　分式析因设计的区组化

8.5　分辨度为III的设计

8.5.1　分辨度为III的设计的构造

8.5.2　折叠分辨度为III的分式设计以分离别名效应

8.5.3　Plackett-Burman 设计

8.6　分辨度为IV和V的设计

8.6.1　分辨度为IV的设计

8.6.2　分辨度为IV的设计的序贯实验

8.6.3　分辨度为V的设计

8.7　超饱和设计

8.8　小结

第 8 章补充材料

S8.1　用于分析分式析因设计的 Yates 法

S8.2　分式析因设计与其他设计的别名结构

S8.3　关于分式析因设计的折叠与部分折叠的更多介绍

8.1　引　　言

当 2^k 析因设计中的因子个数增加时, 设计的一次完全重复所需做的试验次数迅速增大, 以至超出大多数实验者所拥有的资源. 例如, 一个 2^6 设计的完全重复需要做 64 次试验. 在此设计中, 63 个自由度中仅有 6 个与主效应对应, 仅有 15 个自由度对应于二因子交互作用. 其余的 42 个自由度与三因子交互作用及更高阶的交互作用有关.

如果实验者能合理地假定某些高阶交互作用可被忽略, 则关于主效应和低阶交互作用的信息就可只做完全析因实验的一部分而求得. 此类**分式析因设计**[①]是在产品设计、过程设计以及过程改进方面得到最广泛应用的一类设计.

分式析因设计主要用于**筛选实验**, 此类实验是要在众多因子中识别出有 (如果有的话) 大效应的那些因子. 筛选实验通常在项目的早期阶段进行, 那时, 一开始所考虑的很多因子有可能对响应只有小的效应或没有效应. 那些被识别出的重要因子在随后的实验中将被更为深入的研究.

分式析因设计的成功应用基于下面 3 个关键的思想:

(1) **效应稀疏原理**. 当有很多变量时, 系统或过程很可能被少量几个主效应和低阶交互作用所主宰.

(2) **投影性质**. 分式析因设计可以投影到更强的 (更大的) 由显著性因子的子集组成的设计中去.

① 也称为析因设计的部分实施. —— 译者注

(3) **序贯实验**. 可以将两个 (或多个) 分式析因设计序贯地组合成一个设计, 用来估计所感兴趣的因子效应和交互作用.

本章将主要讨论这些原理并用几个例子来加以说明.

8.2 2^k 析因设计的 1/2 分式设计

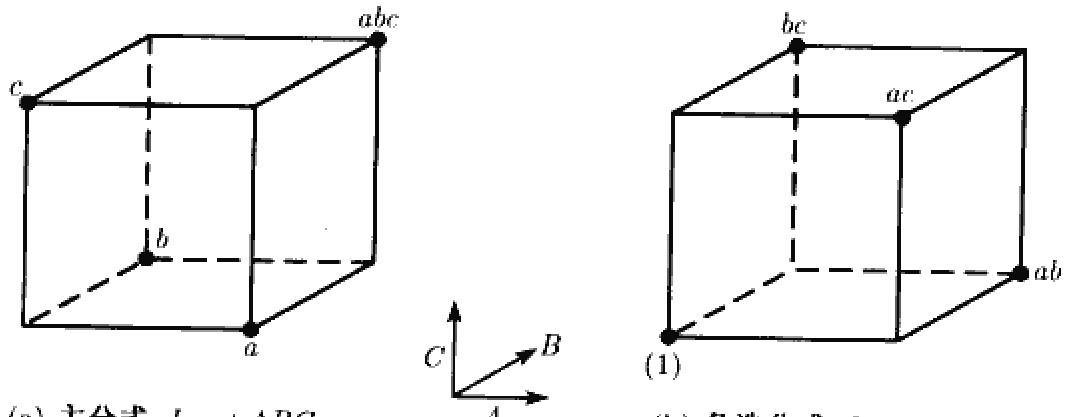
8.2.1 定义与基本原理

考虑有 3 个因子. 每个因子有 2 个水平三因子的设计, 但实验者不能承担所有 $2^3=8$ 种处理组合的试验. 不过, 他们能够承担 4 个. 这就提出了一种 2^3 析因设计的 1/2 分式设计. 因为此设计含有 $2^{3-1}=4$ 种处理组合, 2^3 析因设计的 1/2 分式设计常称为 $\mathbf{2^{3-1}}$ **析因设计**.

2^3 设计的加减符号表见表 8.1. 设选取 4 种处理组合 a, b, c, abc 作为我们的 1/2 分式设计. 这些试验如表 8.1 的上半部分和图 8.1a 所示.

表 8.1 2^3 析因设计的加减符号表

处理组合	因子效应							
	I	A	B	C	AB	AC	BC	ABC
a	+	+	−	−	−	−	+	+
b	+	−	+	−	−	+	−	+
c	+	−	−	+	+	−	−	+
abc	+	+	+	+	+	+	+	+
ab	+	+	+	−	+	−	−	−
ac	+	+	−	+	−	+	−	−
bc	+	−	+	+	−	−	+	−
(1)	+	−	−	−	+	+	+	−



(a) 主分式, $I=+ABC$　　(b) 备选分式, $I=-ABC$

图 8.1 2^3 析因设计的两个 1/2 分式设计

2^{3-1} 设计由选取 ABC 列中带加号的处理组合所组成. 这样一来, ABC 叫做这一特定的分式设计的**生成元** (generator). 有时, 我们称生成元 (比如 ABC) 为一个词 (word). 而且, 单位列 I 全是加号, 因此, 我们称

$$I=ABC$$

为设计的**定义关系** (defining relation). 一般说来, 分式析因设计的定义关系总是与单位列 I 相等的各列构成的集合.

2^{3-1} 设计处理组合有 3 个自由度, 可用来估计主效应. 根据表 8.1, 用来估计 A, B, C 主效应的观测值的线性组合是

$$[A] = \frac{1}{2}(a - b - c + abc), \quad [B] = \frac{1}{2}(-a + b - c + abc), \quad [C] = \frac{1}{2}(-a - b + c + abc)$$

其中记号 $[A], [B], [C]$ 是用来标记与主效应相关的线性组合. 容易证明: 用于估计二因子交互作用的观测值的线性组合是

$$[BC] = \frac{1}{2}(a - b - c + abc), \quad [AC] = \frac{1}{2}(-a + b - c + abc), \quad [AB] = \frac{1}{2}(-a - b + c + abc)$$

于是, $[A] = [BC], [B] = [AC], [C] = [AB]$, 因此, 不可能区别 A 和 BC, B 和 AC, C 和 AB. 事实上, 当我们估计 A, B, C 时, 实际上是估计 $A + BC, B + AC, C + AB$. 具有这一性质的两个或多个效应叫做**别名** (alias). 在我们的例子中, A 和 BC 是别名, B 和 AC 是别名, C 和 AB 是别名. 用符号记为 $[A] \to A + BC, [B] \to B + AC, [C] \to C + AB$.

此设计的别名结构很容易用定义关系 $I = ABC$ 来确定. 任一列 (或效应) 乘以定义关系就得出那一列 (或效应) 的别名. 在我们的例子中, 如 A 的别名可以这样得出:

$$A \cdot I = A \cdot ABC = A^2BC$$

因为任一列的平方恰是单位 I, 所以

$$A = BC$$

同理, 求得 B 和 C 的别名分别为

$$B \cdot I = B \cdot ABC, \qquad B = AB^2C = AC$$

以及

$$C \cdot I = C \cdot ABC, \qquad C = ABC^2 = AB$$

具有 $I = +ABC$ 的 1/2 分式, 通常叫做**主分式** (principal fraction).

现在假定我们选取了另一个 1/2 分式, 即, 表 8.1 中 ABC 列取减号的那些处理组合. 这一**备选的** (alternate) 或**互补的** (complementary)1/2 分式设计 (由试验 (1), ab, ac, bc 组成) 如图 8.1b 所示. 此设计的定义关系是

$$I = -ABC$$

观测值的线性组合, 即 $[A]', [B]', [C]'$, 由备选分式设计给出:

$$[A]' \to A - BC, \qquad [B]' \to B - AC, \qquad [C]' \to C - AB$$

于是, 当我们用这一指定的分式设计来估计 A, B, C 时, 实际上是估计 $A - BC, B - AC, C - AB$.

在实践中, 实际上用哪一分式设计是无关重要的. 两个分式设计属于同一**族** (family), 也就是说, 两个 1/2 分式设计形成一个完全的 2^3 析因设计. 参考图 8.1 的 a 部分和 b 部分就容易看出这一点.

假定做了一个 2^3 析因设计的 1/2 分式设计的试验之后, 另一个也做了. 于是, 所有与 2^3 设计有关的 8 个试验现在都有了. 现在, 将这 8 个试验作为每区组有 4 个试验的二区组 2^3 析因

设计来分析, 就可求得所有效应的分离别名) 的估计量. 这一点, 也可以利用两个分式设计所得的效应的线性组合相加或相减求得. 例如, 考虑 $[A] \to A + BC$ 和 $[A]' \to A - BC$, 可得

$$\frac{1}{2}([A] + [A]') = \frac{1}{2}(A + BC + A - BC) \to A$$
$$\frac{1}{2}([A] - [A]') = \frac{1}{2}(A + BC - A + BC) \to BC$$

于是, 对所有的 3 对线性组合, 可以求得:

i	由 $\frac{1}{2}([i] + [i]')$	由 $\frac{1}{2}([i] - [i]')$
A	A	BC
B	B	AC
C	C	AB

8.2.2　设计分辨度

前面的 2^{3-1} 设计叫做**分辨度为 Ⅲ 的设计** (resolution III design). 在这样的设计中, 主效应的别名为二因子交互作用. 一设计叫做分辨度为 R 的, 若不存在 p 因子效应的别名为另一含有少于 $R-p$ 个因子的效应. 通常用罗马数字下标表示设计的分辨度, 于是, 有定义关系 $I = ABC$ (或 $I = -ABC$) 的 2^3 析因设计的 1/2 分式设计记为 2_{III}^{3-1} 设计.

分辨度为III, IV, V 的设计特别重要. 这些设计的定义及例子如下.

(1) **分辨度为 Ⅲ 的设计**. 其中主效应的别名不能为任一另外的主效应, 但主效应的别名可以为二因子交互作用, 而且一些二因子交互作用可相互为别名. 表 8.1 的 2^{3-1} 设计是分辨度为III的设计 (2_{III}^{3-1}).

(2) **分辨度为 Ⅳ 的设计**. 其中主效应的别名不能为任一另外的主效应或任一二因子交互作用, 但二因子交互作用的别名可为其他的二因子交互作用. 有 $I = ABCD$ 的 2^{4-1} 设计是分辨度为IV的设计 (2_{IV}^{4-1}).

(3) **分辨度为 V 的设计**. 其中主效应或二因子交互作用的别名不能为任一另外的主效应或二因子交互作用, 但二因子交互作用的别名可以为三因子交互作用. 有 $I = ABCDE$ 的 2^{5-1} 设计是分辨度为 V 的设计 (2_{V}^{5-1}).

一般说来, 二水平分式析因设计的分辨度等于定义关系中任一词的字母个数的最小数. 因此, 可以把前面的设计类型分别叫做 3 字母的、4 字母的、5 字母的设计. 通常, 我们喜欢用这样的分式设计: 在所要求的分式程度下, 它具有最高可能的分辨度. 分辨度越高, 对假定条件的限制就越少, 在为了获得对数据的唯一解释而去考虑可以忽略哪些交互作用时就会用到这些假定条件.

8.2.3　1/2 分式设计的构造与分析

一个 2^k 析因设计的 1/2 分式设计的最高分辨度构造法如下. 先写出由完全 2^{k-1} 析因设计的试验组成的**基本设计** (basic design), 然后加进用它们的加号水平和减号水平来判定取加号或减号的最高阶交互作用 $ABC\cdots(K-1)$ 作为第 K 个因子. 因此, 先写出完全的 2^2 析因设计作为基本设计, 然后令因子 C 等于交互作用 AB, 就求得 2^{3-1} 分式析因设计. 备选分式析因

设计令因子 C 等于交互作用 $-AB$ 即得. 表 8.2 说明了这一方法. 基本设计的试验次数 (行数) 正好, 但缺少一列. 用生成元 $I = ABC\cdots K$ 解出缺少的列 (K), 所以 $K = ABC\cdots(K-1)$ 是第 K 个因子, 其水平由各行中各加号与减号的乘积来确定.

表 8.2 2^3 析因设计的两个 1/2 分式设计

试验	完全的 2^2 析因设计 (基本设计)		$2^{3-1}_{\text{III}}, I = ABC$			$2^{3-1}_{\text{III}}, I = -ABC$		
	A	B	A	B	$C = AB$	A	B	$C = -AB$
1	$-$	$-$	$-$	$-$	$+$	$-$	$-$	$-$
2	$+$	$-$	$+$	$-$	$-$	$+$	$-$	$+$
3	$-$	$+$	$-$	$+$	$-$	$-$	$+$	$+$
4	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$-$

任一交互作用效应都可用来生成第 K 个因子的那一列. 不过, 采用任一不同于 $ABC\cdots(K-1)$ 的其他效应, 都得不到最高可能分辨度的设计.

另一个构造 1/2 分式析因设计的方法, 是把试验分解为与最高阶交互作用 $ABC\cdots K$ 相混的两个区组. 每一区组是最高分辨度的 2^{k-1} 分式析因设计.

1. 分式析因设计在析因设计中的投影

任一分辨度为 R 的分式析因设计包含任一 $R-1$ 个因子的子集的完全析因设计 (可能有重复的). 这是一个重要的而且有用的概念. 例如, 实验者有几个可能感兴趣的因子, 但相信只有其中的 $R-1$ 个有重要的效应, 于是, 取分辨度为 R 的分式析因设计是恰当的选择. 如果实验者是正确的, 则分辨度为 R 的分式析因设计将投影到 $R-1$ 个显著性因子的完全析因设计上. 图 8.2 对 2^{3-1}_{III} 设计说明了这一过程, 它投影到每个二因子子集的 2^2 设计上.

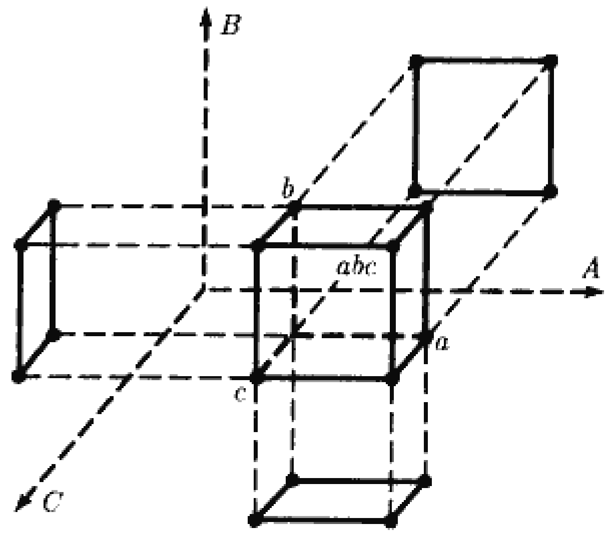


图 8.2 2^{3-1}_{III} 设计投影到 3 个 2^2 设计

因为 2^k 析因设计的 1/2 分式设计最大可能的分辨度是 $R = k$, 每个 2^{k-1} 分式设计可投影到原来 k 个因子的任意 $(k-1)$ 个构成的完全析因设计上. 还有, 2^{k-1} 分式设计又可投影到两次重复的任一 $k-2$ 个因子子集构成的完全析因设计, 4 次重复的任一 $k-3$ 个因子子集构成的完全析因设计, 等等.

例 8.1 考虑例 6.2 的渗透率实验. 原来的设计是单次重复的 2^4 设计, 如表 6.10 所示. 在这个例子中, 我们发现, 主效应 A, C, D 以及交互作用 AC 和 AD 都不等于零. 现在, 重新回到这个实验, 用 2^4 析因设计的 1/2 分式设计来替代原来的完全析因设计, 模拟这一实验, 看看将会出现什么情况.

我们用以 $I = ABCD$ 为生成元的 2^{4-1} 分式析因设计, 最高可能的分辨度为Ⅳ. 要构造出这一设计, 先写出基本设计 (它是 2^3 析因设计), 如表 8.3 头 3 列所示. 基本设计有必需的试验次数 (8 次). 但只有 3 列 (因子). 要求出第 4 个因子的水平, 用 $I = ABCD$ 来求出 D, 即 $D = ABC$. 这样, D 在每次试验所取的水平就是列 A, B, C 的加减符号的乘积. 表 8.3 说明了这种方法. 因为生成元 $ABCD$ 是正的, 这一

2^{4-1}_{IV} 设计就是主分式设计. 图 8.3 是此设计的图解.

表 8.3 定义关系 $I = ABCD$ 的 2^{4-1}_{IV} 设计

试验	基本设计			$D=ABC$	处理组合	渗透率
	A	B	C			
1	−	−	−	−	(1)	45
2	+	−	−	+	ad	100
3	−	+	−	+	bd	45
4	+	+	−	−	ab	65
5	−	−	+	+	cd	75
6	+	−	+	−	ac	60
7	−	+	+	−	bc	80
8	+	+	+	+	$abcd$	96

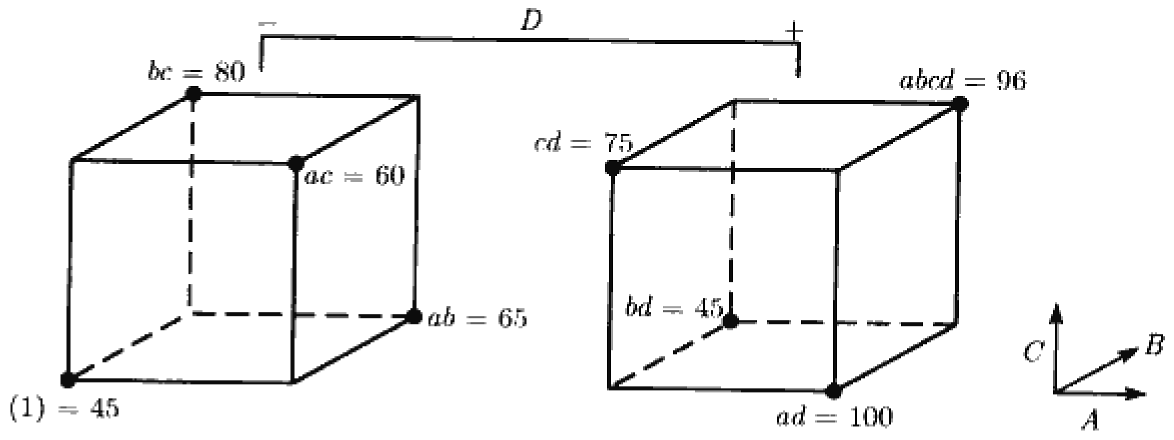


图 8.3 例 8.1 渗透率实验的 2^{4-1}_{IV} 设计

利用定义关系, 每一主效应的别名是三因子交互作用. 也就是说, $A = A^2BCD = BCD$, $B = AB^2CD = ACD$, $C = ABC^2D = ABD$, $D = ABCD^2 = ABC$. 而且, 每个二因子交互作用的别名是另一个二因子交互作用. 这些别名关系是 $AB = CD, AC = BD, BC = AD$. 4 个主效应加上 3 个二因子交互作用别名对, 总共占用了此设计的 7 个自由度.

此时, 依随机顺序做这 8 个试验. 因为已经做了完全 2^4 设计的实验, 所以我们简单地从例 6.2 中选取 8 个观测到的渗透率与 2^{4-1}_{IV} 设计的实验相对应. 这些观测值列于表 8.3 最后一列, 也标在图 8.3 中.

由此 2^{4-1}_{IV} 设计求得的效应估计量见表 8.4. 为说明计算方法, 与 A 效应有关的观测值的线性组合是

$$[A] = \frac{1}{4}(-45+100-45+65-75+60-80+96) = 19.00 \to A + BCD$$

而对于 AB 效应, 求得

$$[AB] = \frac{1}{4}(45-100-45+65+75-60-80+96) = -1.00 \to AB + CD$$

审核表 8.4 中的信息, 有理由认为主效应 A, C, D 较大. 而且, 若 A, C, D 是重要的主效应, 因为交互作用 AC 与 AD 也是显著的, 则两个交互作用别名 $AC + BD$ 与 $AD + BC$ 自然会有大的效应. 这里运用了**Ockham 剃刀原则** (William Ockham 所称), 即当碰到一个现象有多个不同的可能解释时, 最简单的解释通常是正确的. 可以看出上述解释与例 6.2 中完全 2^4 设计的分析所得的结论一致.

因为因子 B 不显著, 对它可以不加考虑. 因此, 可以将这 2^{4-1}_{IV} 设计投影到关于因子 A, C, D 的单次重复的 2^3 设计上, 如图 8.4 所示. 对此立方图的检查使我们对上面所得到的结论更加放心. 当温度 (A) 处于低水平时, 浓度 (C) 有较大的正效应; 而当温度处于高水平时, 浓度有很小的效应. 这很可能是由 AC 的交互作用引起的. 而且, 当温度处于低水平时, 搅拌速度 (D) 可被忽略; 而当温度处于高水平时, 搅拌速度有较大的正效应. 这很可能是由前面识别出的 AD 交互作用引起的.

表 8.4 例 8.1 的效应与别名的估计量 [a]

估 计	别名结构
$[A] = 19.00$	$[A] \to A + BCD$
$[B] = 1.50$	$[B] \to B + ACD$
$[C] = 14.00$	$[C] \to C + ABD$
$[D] = 16.50$	$[D] \to D + ABC$
$[AB] = -1.00$	$[AB] \to AB + CD$
$[AC] = -18.50$	$[AC] \to AC + BD$
$[AD] = 19.00$	$[AD] \to AD + BC$

a 显著的效应是 A, C, D, AC, AD.

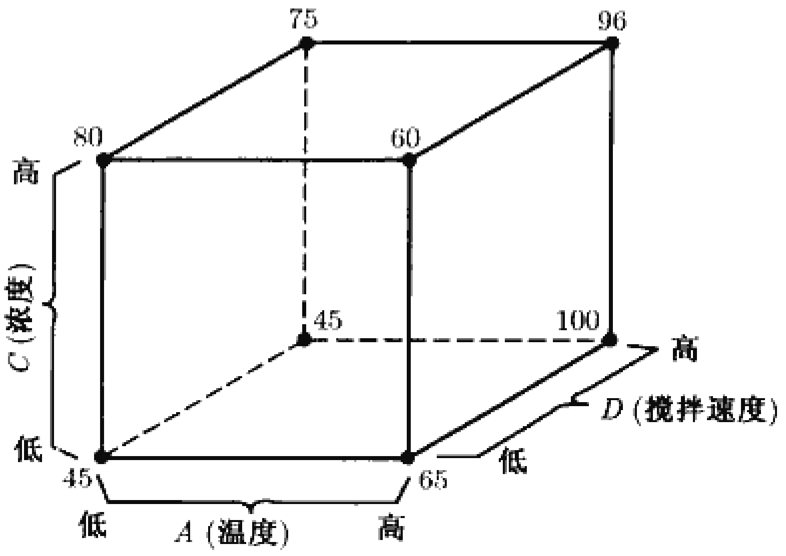


图 8.4 例 8.1 的 2_{IV}^{4-1} 设计在 A, C, D 的 2^3 设计上的投影

基于上述分析, 可以得到用于预测实验区域内的渗透率的模型. 此模型为

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_{13} x_1 x_3 + \hat{\beta}_{14} x_1 x_4$$

其中 x_1, x_3, x_4 是表示 A, C, D 的规范变量 $(-1 \leqslant x_i \leqslant +1)$, 像前面一样, $\hat{\beta}$ 是由效应估计得到的回归系数. 因此, 预测方程为

$$\hat{y} = 70.75 + \left(\frac{19.00}{2}\right) x_1 + \left(\frac{14.00}{2}\right) x_3 + \left(\frac{16.50}{2}\right) x_4 + \left(\frac{-18.50}{2}\right) x_1 x_3 + \left(\frac{19.00}{2}\right) x_1 x_4$$

记住截距项 $\hat{\beta}_0$ 是实验的所有 8 次试验的响应值的平均值. 此模型与例 6.2 中根据 2^k 完全析因设计所得的模型非常相似.

例 8.2 用来改进生产过程的 2^{5-1} 分式设计

在一集成电路生产线上用 2^{5-1} 设计来研究 5 个因子, 以改进生产量. 5 个因子是: A= 孔径设定 (小, 大), B= 曝光时间 (低于额定的 20%, 高于额定的 20%), C= 冲洗时间 (30 秒, 45 秒), D= 掩膜尺寸 (小, 大), E = 蚀刻时间 (14.5 分钟, 15.5 分钟). 2^{5-1} 分式设计的构造如表 8.5 所示. 构造此设计的方法是, 先写出有 16 个试验的基本设计 (A, B, C, D 的一个 2^4 设计), 选取 $ABCDE$ 为生成元. 然后设定第 5 个因子 $E = ABCD$ 的水平. 图 8.5 给出了这一设计的图形表示.

此设计的定义关系是 $I = ABCDE$. 因此, 每一主效应的别名是一个四因子交互作用 (例如, $[A] \to A + BCDE$), 每一个二因子交互作用的别名是一个三因子交互作用 (例如, $[AB] \to AB + CDE$). 这样, 该设计的分辨度为 V. 我们希望这一 2^{5-1} 设计能很好地提供关于主效应和二因子交互作用的信息.

表 8.6 列出了这一实验的 15 个效应的估计量、平方和与回归系数. 图 8.6 画出了这一实验的效应估计量的正态概率图. 主效应 A, B, C 以及 AB 交互作用较大. 因为别名关系, 这些效应实际上是 $A + BCDE$,

表 8.5　例 8.2 的 2_{V}^{5-1} 设计

试验	基本设计				$E=ABCD$	处理组合	产量
	A	B	C	D			
1	−	−	−	−	+	e	8
2	+	−	−	−	−	a	9
3	−	+	−	−	−	b	34
4	+	+	−	−	+	abe	52
5	−	−	+	−	−	c	16
6	+	−	+	−	+	ace	22
7	−	+	+	−	+	bce	45
8	+	+	+	−	−	abc	60
9	−	−	−	+	−	d	6
10	+	−	−	+	+	ade	10
11	−	+	−	+	+	bde	30
12	+	+	−	+	−	abd	50
13	−	−	+	+	+	cde	15
14	+	−	+	+	−	acd	21
15	−	+	+	+	−	bcd	44
16	+	+	+	+	+	$abcde$	63

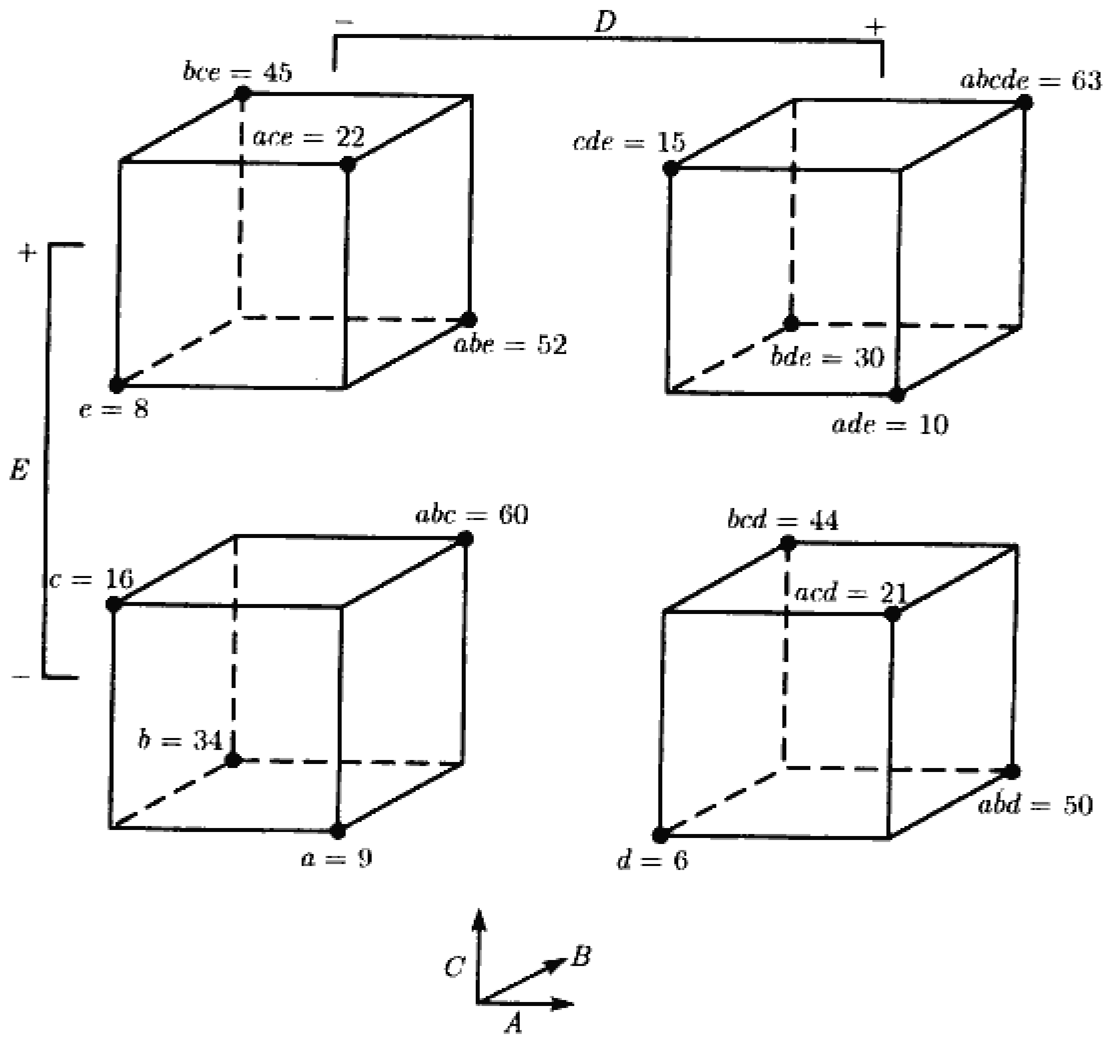


图 8.5　例 8.2 的 2_{V}^{5-1} 设计

表 8.6　例 8.2 的效应、回归系数与平方和

变量	名称	−1 水平	+1 水平
A	孔径设定	小	大
B	曝光时间	−20%	+20%
C	冲洗时间	30 s	40 s
D	掩膜尺寸	小	大
E	蚀刻时间	14.5 min	15.5 min

(续)

变量	回归系数	效应估计	平方和
总平均值	30.312 5		
A	5.562 5	11.125 0	495.062
B	16.937 5	33.875 0	4 590.062
C	5.437 5	10.875 0	473.062
D	−0.437 5	−0.875 0	3.063
E	0.312 5	−0.625 0	1.563
AB	3.437 5	6.875 0	189.063
AC	0.187 5	0.375 0	0.563
AD	0.562 5	1.125 0	5.063
AE	0.562 5	1.125 0	5.063
BC	0.312 5	0.625 0	1.563
BD	−0.062 5	−0.125 0	0.063
BE	−0.062 5	−0.125 0	0.063
CD	0.437 5	0.875 0	3.063
CE	0.187 5	0.375 0	0.563
DE	−0.687 5	−1.375 0	7.563

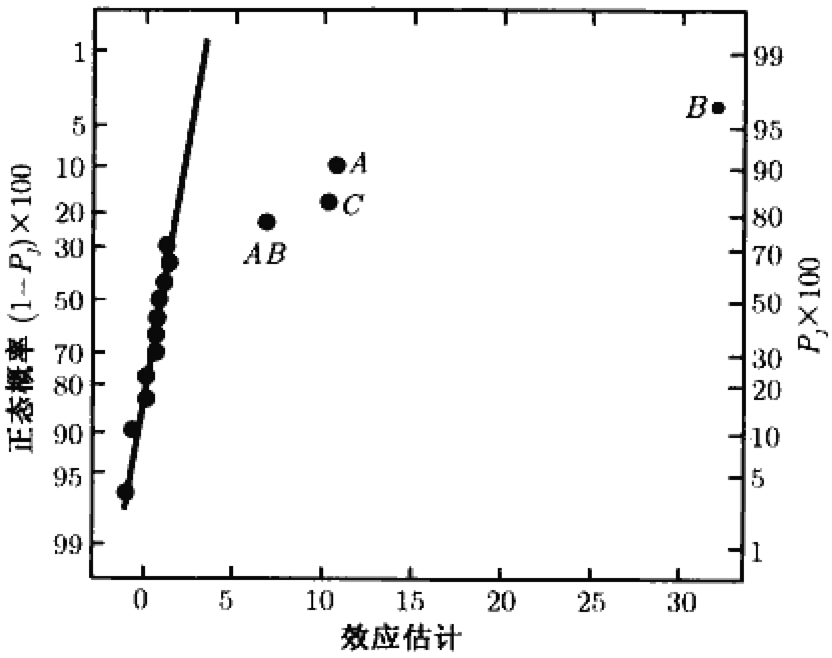


图 8.6 例 8.2 效应的正态概率图

$B+ACDE$, $C+ABDE$, $AB+CDE$. 不过, 因为三因子交互作用和更高的交互作用可被忽略这一点看起来挺有道理, 因此, 可以认为只有 A, B, C 和 AB 是重要的效应.

表 8.7 是此实验的方差分析表. 模型平方和是 $SS_{模型}=SS_A+SS_B+SS_C+SS_{AB}=5\ 747.25$, 它占了产量总变异性的 99%以上. 图 8.7 是残差的正态概率图. 图 8.8 是残差与预测值的关系图. 这两张图都是令人满意的.

表 8.7 例 8.2 的方差分析

方差来源	平方和	自由度	均　方	F_0	P 值
A(孔径设定)	495.062 5	1	495.062 5	193.20	< 0.000 1
B(曝光时间)	4 590.062 5	1	4 590.062 5	1 791.24	< 0.000 1
C(冲洗时间)	473.062 5	1	473.062 5	184.61	< 0.000 1
AB	189.062 5	1	189.062 5	73.78	< 0.000 1
误差	28.187 5	11	2.562 5		
总和	5 775.437 5	15			

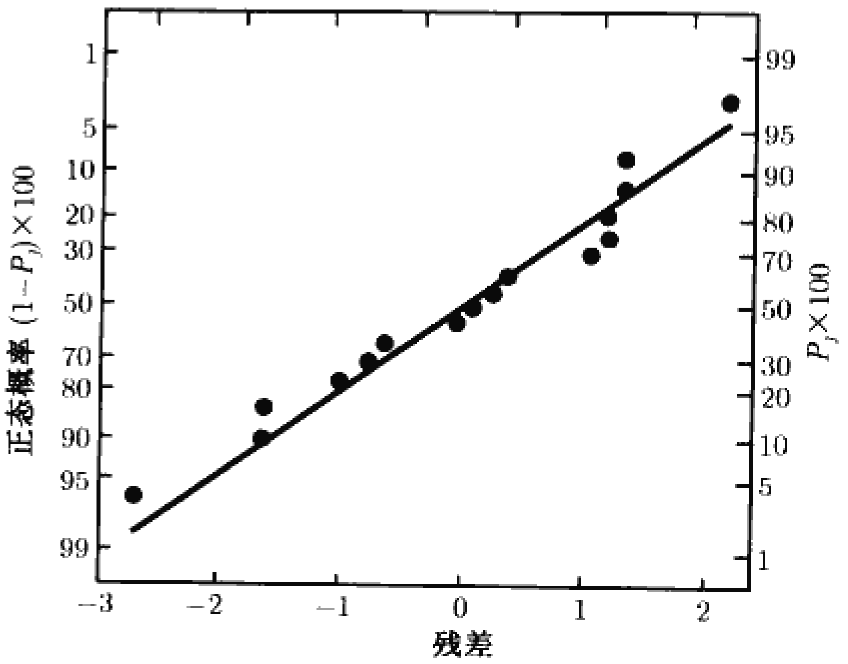


图 8.7　例 8.2 残差的正态概率图

3 个因子 A, B, C 有较大的正效应. 图 8.9 画出了 AB 或孔径设定–曝光时间交互作用的图形. 这一图形表明, 当 A 与 B 都处于高水平时产量较高.

2^{5-1} 设计可压缩为由原来 5 个因子中的任意 3 个构成的 2^3 设计的两次重复 (参见图 8.5). 图 8.10 是因子 A, B, C 的立方体图, 在 8 个角点上注上了平均产量. 审查此立方图, 显然可见, 最高的产量在 A, B, C 都处于高水平时达到. 因子 D 和 E 对平均产量有较小的效应. 因此, 它们的取值大小可从优化其他目标 (例如成本) 考虑.

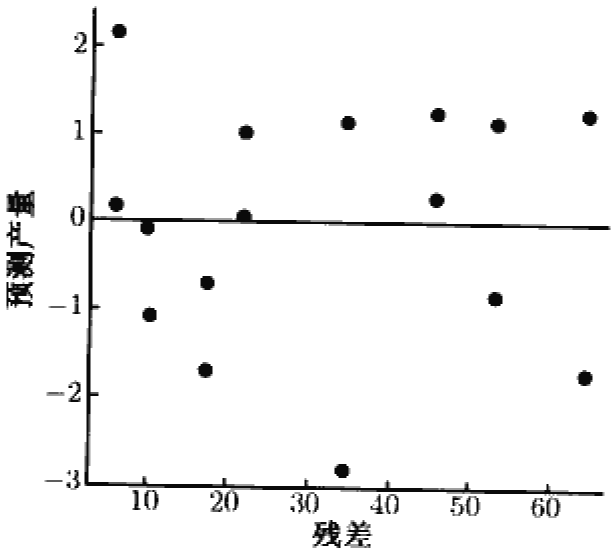


图 8.8　例 8.2 残差与预测值的关系图

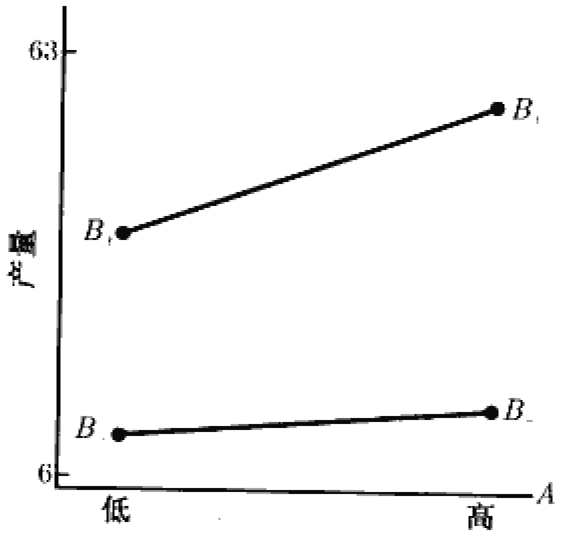


图 8.9　例 8.2 孔径设定–曝光时间的交互作用

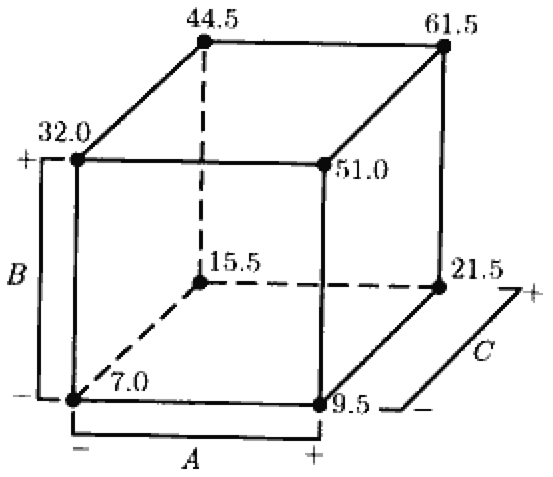


图 8.10　例 8.2 的 2_{V}^{5-1} 设计在因子 A, B, C 构成的 2^3 设计的两次重复中的投影

2. 分式析因设计的试验顺序

用分式析因设计通常使得实验既省钱效率又高, 特别是在实验可以序贯地进行时. 例如, 假定要研究 $k=4$ 个因子 ($2^4=16$ 个试验), 则我们经常喜欢做 2_{IV}^{4-1} 分式设计 (8 个试验), 分析其结果, 然后决定下一次最好做哪些试验. 当有必要去弄清模糊不清的情况时, 就可再做备选的分式设计实验, 完成 2^4 设计的实验. 当用这种方法来完成实验设计时, 两个 1/2 分式设计实验代表将最高阶交互作用与区组相混 (此处是混

杂了 $ABCD$) 的完全混区设计的两个区组. 这样一来, 序贯实验的结果仅仅损失了最高阶交互作用的信息. 另外, 在大多数情况下, 从 1/2 分式析因设计实验中, 我们会非常清楚进行下一阶段的实验时, 应该加进哪些因子, 去掉哪些因子, 改变哪些响应变量, 或者改变某些因子的取值范围等. 这些可能情形以图形方式列在图 8.11 中.

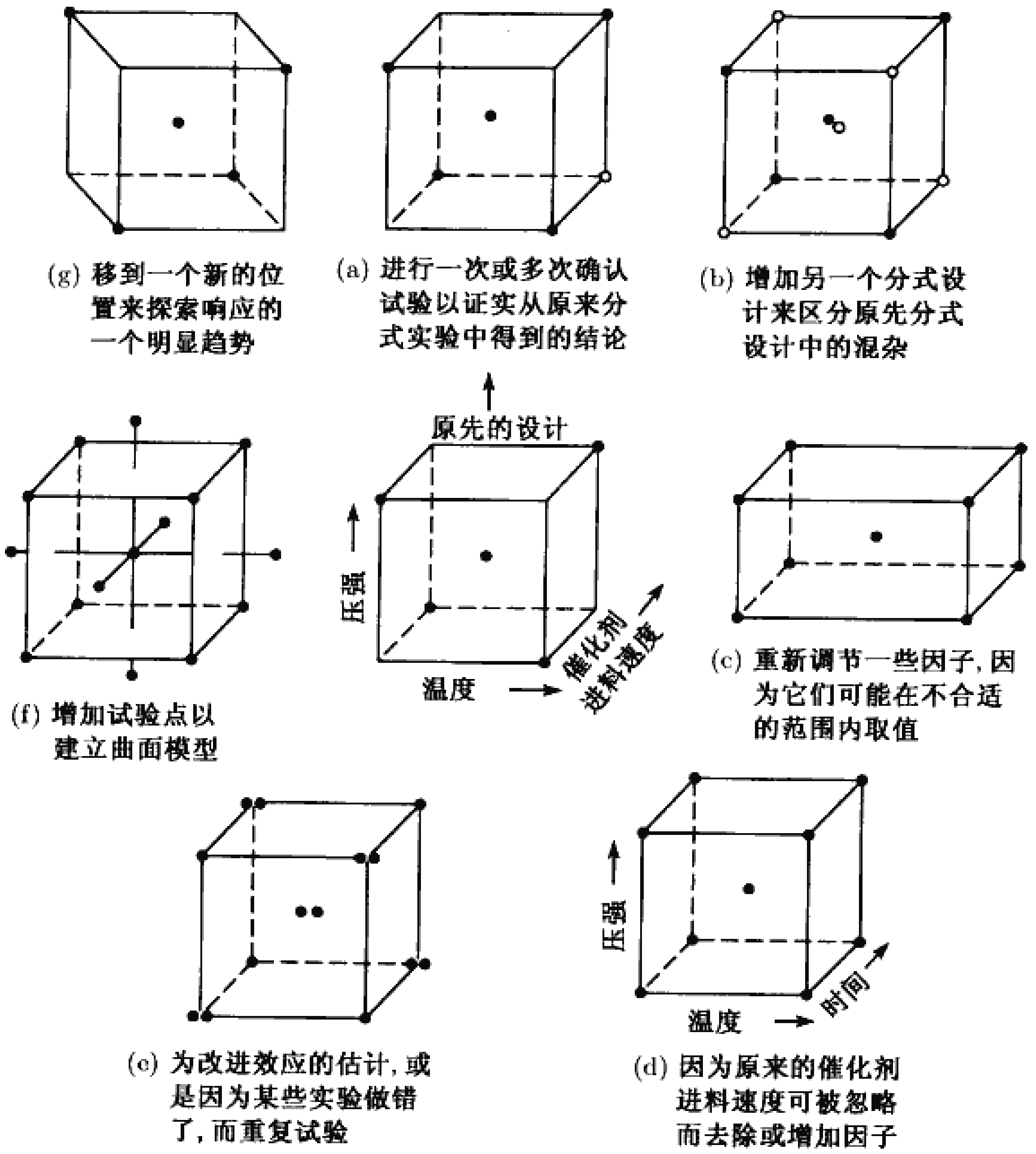


图 8.11 分式析因实验后追加实验的几种可能性 [取自 Box(1992~1993), 经出版商允许]

例 8.3 重新考虑例 8.1 中的实验. 用 2_{IV}^{4-1} 设计初步识别出 3 个大的主效应: A, C, D. 有两个大的效应与二因子交互作用有关: $AC+BD$ 和 $AD+BC$. 在例 8.2 中, 根据主效应 B 可以忽略, 初步认为重要的交互作用是 AC 和 AD. 有时, 实验者具备某些工序知识, 从而方便判别交互作用中哪个可能更重要. 不过, 完成由 $I=-ABCD$ 确定的备选分式设计后, 总能找出显著的交互作用. 以下直接列出此设计与响应:

这些效应估计量 (以及它们的别名) 由上述备选分式设计算得:

$$[A]'=24.25 \rightarrow A-BCD$$
$$[B]'=4.75 \rightarrow B-ACD$$
$$[C]'=5.75 \rightarrow C-ABD$$
$$[D]'=12.75 \rightarrow D-ABC$$
$$[AB]'=1.25 \rightarrow AB-CD$$
$$[AC]'=-17.75 \rightarrow AC-BD$$
$$[AD]'=14.25 \rightarrow AD-BC$$

试验	基本设计			$D=-ABC$	处理组合	渗透率
	A	B	C			
1	−	−	−	+	d	43
2	+	−	−	−	a	71
3	−	+	−	−	b	48
4	+	+	−	+	abd	104
5	−	−	+	−	c	68
6	+	−	+	+	acd	86
7	−	+	+	+	bcd	70
8	+	+	+	−	abc	65

这些估计量和由原来的 1/2 分式设计所得的估计量组合起来就得到下列效应估计量:

i	由 $\frac{1}{2}([i]+[i]')$	由 $\frac{1}{2}([i]-[i]')$
A	$21.63 \to A$	$-2.63 \to BCD$
B	$3.13 \to B$	$-1.63 \to ACD$
C	$9.88 \to C$	$4.13 \to ABD$
D	$14.63 \to D$	$1.88 \to ABC$
AB	$0.13 \to AB$	$-1.13 \to CD$
AC	$-18.13 \to AC$	$-0.38 \to BD$
AD	$16.63 \to AD$	$2.38 \to BC$

这些估计量与原先例 6.2 中作为单次重复的 2^4 析因设计的数据分析结果完全一致. 显然, 大的交互作用是 AC 和 AD.

把备选分式设计添加到主分式设计可以看作是一种**确认实验** (confirmation experiment), 这是因为它提供的信息使我们强化了关于二因子交互作用效应的最初结论. 8.5 节与 8.6 节中还将研究用于区分交互作用的组合分式析因设计的其他一些方面.

确认实验并不需要这样复杂. 一种非常简单的确认实验是: 为了利用模型方程去预测在设计空间的某个感兴趣点 (并非该设计中的点) 处的响应值, 那么就以该处理组合实际进行试验 (也许重复几次), 然后比较响应的实际值与预测值就行了. 它们相当接近则表明, 分式析因的解释是正确的, 反之, 若存在很大不同则意味着解释存在问题. 由此可以看到, 附加实验可用于区分混淆.

以例 8.1 中的 2^{4-1} 分式析因设计为例. 实验者想要找出响应变量渗透率较高的一组条件, 但是要求甲醛浓度 (因子 C) 比较低. 这表明因子 A 和 D 应该取高水平, 而因子 C 应该取低水平. 检查图 8.3, 可以看到, 当 B 取低水平时, 在分式析因设计中该处理组合进行了试验, 得到响应观测值为 100. 但是 B 取高水平时的该处理组合不在原先设计中, 故此处理组合应该是合理的确认试验. 在因子 A, B, D 取高水平而因子 C 取低水平时, 由例 8.1 的模型方程可以算得预测响应如下:

$$\begin{aligned}\hat{y} =&70.75+\left(\frac{19.00}{2}\right)x_1+\left(\frac{14.00}{2}\right)x_3+\left(\frac{16.50}{2}\right)x_4+\left(\frac{-18.50}{2}\right)x_1x_3+\left(\frac{19.00}{2}\right)x_1x_4\\ =&70.75+\left(\frac{19.00}{2}\right)(1)+\left(\frac{14.00}{2}\right)(-1)+\left(\frac{16.50}{2}\right)(1)+\left(\frac{-18.50}{2}\right)(1)(-1)+\left(\frac{19.00}{2}\right)(1)(1)\\ =&100.25\end{aligned}$$

该处理组合处的响应观测值是 104(参见图 6.10, 其中给出了完全的 2^4 析因设计的各种响应数

据). 因为渗透率的观测值与预测值非常接近, 所以这是一个成功的确认实验. 这也表明分式析因实验的解释是正确的.

有时一个确认试验的预测值与观测值并不是这么接近, 这就有必要回答两个值是否足够接近, 从而有理由认为分式设计的解释是正确的. 回答这个问题的一种方法是: 构造该确认试验的未来响应值的**预测区间** (prediction interval), 然后看实际观测值是否落在预测区间中. 在 10.6 节介绍回归模型的预测区间时, 我们将用这个例子来说明这一点.

8.3 2^k 析因设计的 1/4 分式设计

对中等大的因子个数, 常用 2^k 设计的较小的分式设计. 考虑 2^k 设计的 1/4 分式设计. 此设计包含 2^{k-2} 个试验, 通常叫做 **2^{k-2} 分式析因设计**.

2^{k-2} 设计的构造方法是, 首先写出 $k-2$ 个因子的完全析因设计的试验组成的**基本设计**, 然后再加上两列, 这两列由开头的 $k-2$ 个因子的恰当交互作用组成的. 这样一来, 2^{k-2} 设计的 1/4 分式设计有两个生成元. 如果 P 和 Q 代表选取的生成元, 则, $I = P$ 和 $I = Q$ 叫做此设计的**生成关系** (generating relation). P 和 Q 的符号 (+ 或者 −) 决定了采用哪一种 1/4 分式. 与**生成元** $\pm P$ 和 $\pm Q$ 的选择有关的所有 4 个分式是同一**族**的成员. P 和 Q 都是正的那个分式是主分式.

所有等于单位列 I 的列组成设计的**完全定义关系**. 这由 P, Q, 以及它们的**广义交互作用** PQ 所组成, 也就是说, 定义关系是 $I = P = Q = PQ$. 称元素 P, Q, PQ 为定义关系词. 任一效应的别名可由效应所在的列乘以定义关系的每个词而得. 显然, 每一效应有 3 个别名. 实验者要谨慎地选择生成元以避免可能重要的效应相互为别名.

作为例子, 考虑 2^{6-2} 设计. 设选取了 $I = ABCE$ 和 $I = BCDF$ 为设计的生成元. 生成元 $ABCE$ 和 $BCDF$ 的广义交互作用是 $ADEF$, 因此, 此设计的完全定义关系是

$$I = ABCE = BCDF = ADEF$$

从而, 此设计分辨度为Ⅳ. 为求得任一效应 (例如 A) 的别名, 将此效应与定义关系的每个词相乘即可. 对于 A, 得

$$A = BCE = ABCDF = DEF$$

容易证明, 每个主效应的别名是三因子交互作用和五因子交互作用, 二因子交互作用的别名是另一个二因子交互作用或更高阶交互作用. 于是, 估计 A 时, 我们实际上是估计 $A + BCE + DEF + ABCDF$. 此设计的全部别名结构见表 8.8. 如果三因子和更高阶的交互作用可被忽略, 则此设计给出了主效应的纯净估计量.

要构造此设计, 首先写出**基本设计**, 它由以 A, B, C, D 为因子的完全 $2^{6-2} = 2^4$ 设计的 16 个试验所组成. 然后加上两个因子 E 和 F, 它们的水平分别由交互作用 ABC 和 BCD 的加减符号所决定. 此方法见表 8.9.

构造此设计的另一方法是, 先使 $ABCE$ 和 $BCDF$ 与区组相混, 导出 2^6 设计的 4 个区组, 然后选取处理组合在 $ABCE$ 和 $BCDF$ 上都为正的区组. 这就得出生成关系为 $I = ABCE$ 和 $I = BCDF$ 的 2^{6-2} 分式析因设计, 因为生成元 $ABCE$ 和 $BCDF$ 都是正的, 这就是主分式.

表 8.8　$I = ABCE = BCDF = ADEF$ 的 2_{IV}^{6-2} 设计的别名结构

$A = BCE = DEF = ABCDF$	$AB = CE = ACDF = BDEF$
$B = ACE = CDF = ABDEF$	$AC = BE = ABDF = CDEF$
$C = ABE = BDF = ACDEF$	$AD = EF = BCDE = ABCF$
$D = BCF = AEF = ABCDE$	$AE = BC = DF = ABCDEF$
$E = ABC = ADF = BCDEF$	$AF = DE = BCEF = ABCD$
$F = BCD = ADE = ABCEF$	$BD = CF = ACDE = ABEF$
	$BF = CD = ACEF = ABDE$
$ABD = CDE = ACF = BEF$	
$ACD = BDE = ABF = CEF$	

表 8.9　生成元 $I = ABCE$ 和 $I = BCDF$ 的 2_{IV}^{6-2} 设计的构造

试验	基本设计			D	$E = ABC$	$E = BCD$
	A	B	C			
1	−	−	−	−	−	−
2	+	−	−	−	+	−
3	−	+	−	−	+	+
4	+	+	−	−	−	+
5	−	−	+	−	+	+
6	+	−	+	−	−	+
7	−	+	+	−	−	−
8	+	+	+	−	+	−
9	−	−	−	+	−	+
10	+	−	−	+	+	+
11	−	+	−	+	+	−
12	+	+	−	+	−	−
13	−	−	+	+	+	−
14	+	−	+	+	−	−
15	−	+	+	+	−	+
16	+	+	+	+	+	+

当然, 此 2^{6-2} 设计还有 3 个**备选分式**设计. 它们是生成关系 $I = ABCE$ 和 $I = -BCDF$, $I = -ABCE$ 和 $I = BCDF$, $I = -ABCE$ 和 $I = -BCDF$ 的分式设计. 这些分式设计容易用表 8.9 所示的方法构成. 例如, 如果要求出 $I = ABCE$ 和 $I = -BCDF$ 的分式设计, 则在表 8.9 的最后一列令 $F = -BCD$, 因子 F 这一列的水平变成

$$+ + - - - - + + - - + + + + - -$$

此备选分式设计的完全定义关系是 $I = ABCE = -BCDF = -ADEF$. 表 8.9 中别名结构的某些符号现在更改了, 例如, A 的别名是 $A = BCE = -DEF = -ABCDF$. 这样一来, 观测值的线性组合 $[A]$ 实际上是估计 $A + BCE - DEF - ABCDF$.

最后, 2_{IV}^{6-2} 分式析因设计可压缩为由 4 个因子构成的任一子集的单次重复的 2^4 设计, 只要这一子集不是定义关系中的一个词即可. 也可以压缩为由 4 个因子构成的任一子集的有重复的 2^4 设计的 1/2 分式设计, 只要那一子集是定义关系中的一个词就行. 如此, 表 8.9 的设计就变成以因子 A, B, C, E, 或 B, C, D, F, 或 A, D, E, F 构成的有两次重复的 2^{4-1} 分式析因设计, 因为那些因子正是定义关系的词. 6 个因子还有另外 12 个组合, 比如 $ABCD$, $ABCF$, 等等, 它

们是此设计在单次重复的 2^4 设计上的投影. 此设计还可压缩为由 6 个因子的任意 3 个组成的有 4 次重复的 2^3 设计, 或由两个因子组成的有 4 次重复的 2^2 设计.

一般说来, 任一 2^{k-2} 分式析因设计可以压缩为由原来的因子的 $r \leqslant k-2$ 个元素的某一子集构成的完全析因设计或分式析因设计. 构成完全析因设计的那些变量的子集不是完全定义关系中的词.

例 8.4 在注塑成型生产线上制造的零件显得过度收缩. 这是在装配时发现的, 寻根探源是由注塑成型工序引起的. 质量改进小组决定用实验设计去研究注塑成型生产过程, 以减小收缩率. 小组决定研究 6 个因子 —— 成型温度 (A), 螺杆转速 (B), 保持时间 (C), 循环时间 (D), 射口大小 (E), 保压压力 (F)— 每个因子取两个水平. 研究的目的是弄清每个因子是怎样影响收缩量的, 因子的交互作用怎么样.

小组决定用表 8.9 的有 16 个试验的二水平分式析因设计. 此设计再度列于表 8.10 中, 在做设计的 16 个试验的每一个试验时, 测试生产零件得到观测的收缩量 (×10) 也一起列在该表中. 表 8.11 列出了该实验的效应估计、平方和以及回归系数.

表 8.10　例 8.4 中注塑成型实验的一个 2_{IV}^{6-2} 设计

试验	基本设计				$E=ABC$	$E=BCD$	收缩量的观测值 (×10)
	A	B	C	D			
1	−	−	−	−	−	−	6
2	+	−	−	−	+	−	10
3	−	+	−	−	+	+	32
4	+	+	−	−	−	+	60
5	−	−	+	−	+	+	4
6	+	−	+	−	−	+	15
7	−	+	+	−	−	−	26
8	+	+	+	−	+	−	60
9	−	−	−	+	−	+	8
10	+	−	−	+	+	+	12
11	−	+	−	+	+	−	34
12	+	+	−	+	−	−	60
13	−	−	+	+	+	−	16
14	+	−	+	+	−	−	5
15	−	+	+	+	−	+	37
16	+	+	+	+	+	+	52

此实验的效应估计值的正态概率图见图 8.12. 较大的效应只有 A (成型温度), B (螺杆转速), 以及 AB 的交互作用. 看一下表 8.8 的别名关系, 暂时接受这些结论看来是合理的. 图 8.13 的 AB 交互作用图表明: 当螺杆转速处于低水平时, 生产过程对温度很不敏感; 当螺杆转速处于高水平时, 对温度十分敏感. 当螺杆转速为低水平, 则不管温度选什么水平, 得到的平均收缩率约为 10%.

据此初步分析, 小组决定成型温度与螺杆转速都用低水平. 这组条件使零件的平均收缩率减至大约 10%. 但是, 零件之间收缩率的变异性仍然是一个问题. 实际上, 用上述的调整方法, 平均收缩率可适当减少. 然而, 零件间收缩率的变异性可能还会在装配上引起问题. 解决问题的一个办法是, 看一下有无因子影响零件收缩率的变异性.

图 8.14 是残差的正态概率图. 此图令人满意. 然后作出残差与每个因子的关系图. 图 8.15 是这些图之一, 残差与因子 C(保持时间) 的关系图. 此图表明, 残差在保持时间处于低水平时远没有处于高水平时那样分散. 这些残差是按常规方法从收缩率的预测模型求得的. 因为预测的收缩率

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 x_2 = 27.312\,5 + 6.937\,5 x_1 + 17.812\,5 x_2 + 5.937\,5 x_1 x_2$$

其中 x_1, x_2, x_1x_2 是规范变量, 分别对应于因子 A, 因子 B, 以及 AB 交互作用, 于是残差为

$$e = y - \hat{y}$$

用来产生残差的回归模型实质上是从数据中消除了 A, B, AB 的**位置效应**，因此，残差包含有未加说明的变异性信息. 图 8.15 表明，变异性中有一定的模式，当保持时间处于低水平时，零件收缩率的变异性较小. (请回忆在第 6 章我们看到的，当位置或均值模型是正确时残差只是反映关于分散效应的信息).

表 8.11 例 8.4 中的效应估计、平方和以及回归系数

变量	名称	−1 水平	+1 水平
A	成型温度	−1.000	1.000
B	螺杆转速	−1.000	1.000
C	保持时间	−1.000	1.000
D	循环时间	−1.000	1.000
E	射口大小	−1.000	1.000
F	保压压力	−1.000	1.000
变量 *	**回归系数**	**效应估计**	**平方和**
总平均值	27.312 5		
A	6.937 5	13.875 0	770.062
B	17.812 5	35.625 0	5 076.562
C	−0.437 5	−0.875 0	3.063
D	0.687 5	1.375 0	7.563
E	0.187 5	0.375 0	0.563
F	0.187 5	0.375 0	0.563
$AB+CE$	5.937 5	11.875 0	564.063
$AC+BE$	−0.812 5	−1.625 0	10.562
$AD+EF$	−2.687 5	−5.375 0	115.562
$AE+BC+DF$	−0.937 5	−1.875 0	14.063
$AF+DE$	0.312 5	0.625 0	1.563
$BD+CF$	−0.062 5	−0.125 0	0.063
$BF+CD$	−0.062 5	−0.125 0	0.063
ABD	0.062 5	0.125 0	0.063
ABF	−2.437 5	−4.875 0	95.063

* 只有主效应与二因子交互作用.

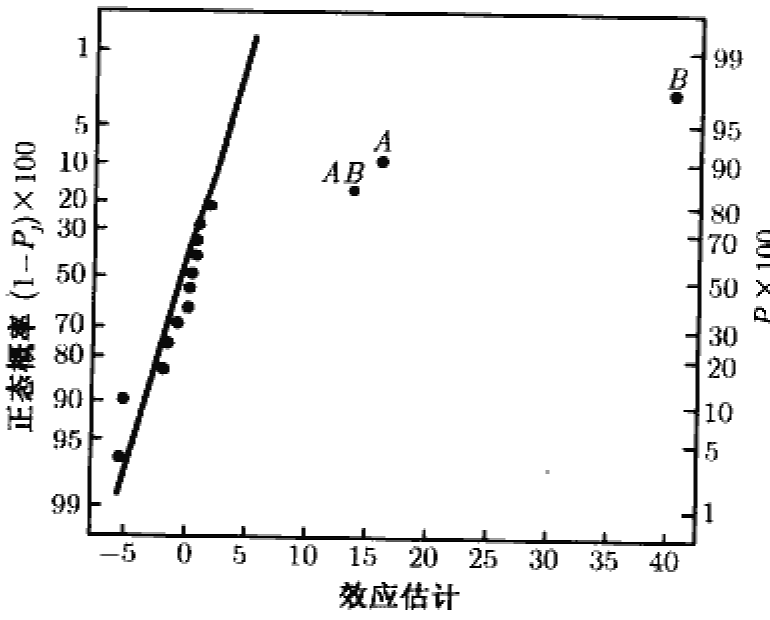


图 8.12 例 8.4 效应的正态概率图

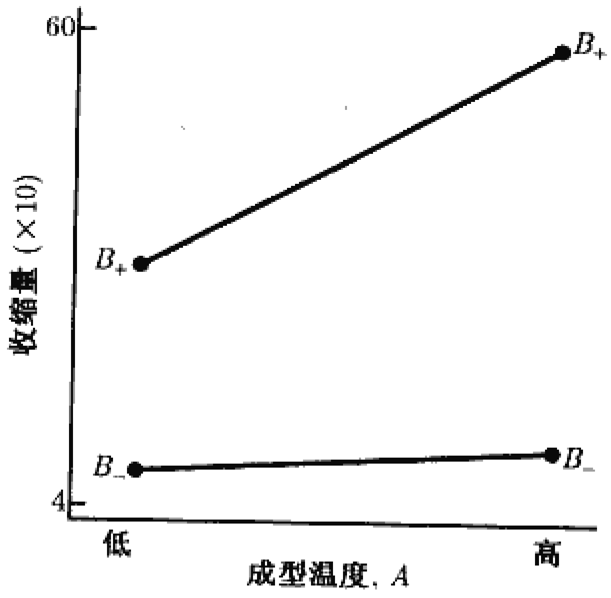


图 8.13 例 8.4 中 AB (成型温度–螺杆转速) 交互作用的图形

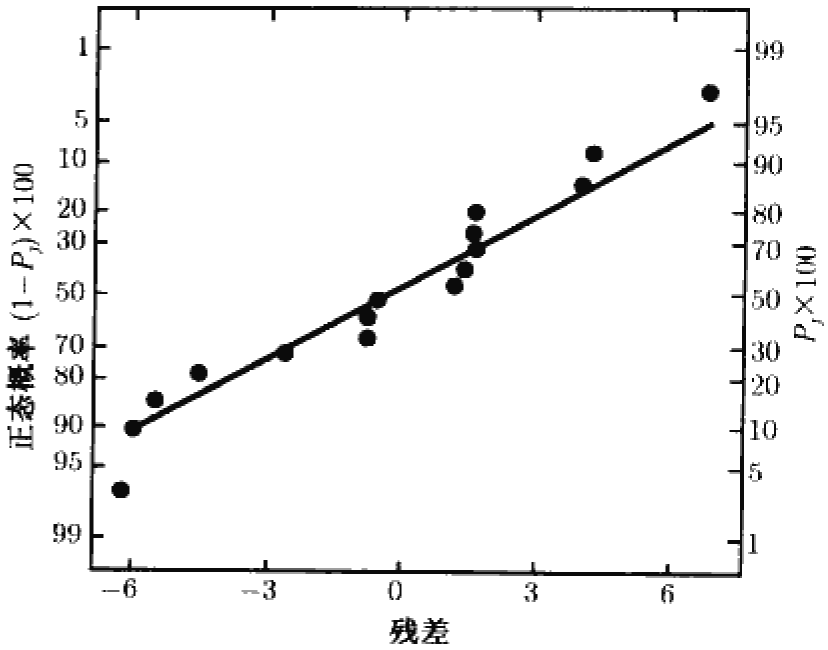


图 8.14 例 8.4 残差的正态概率图

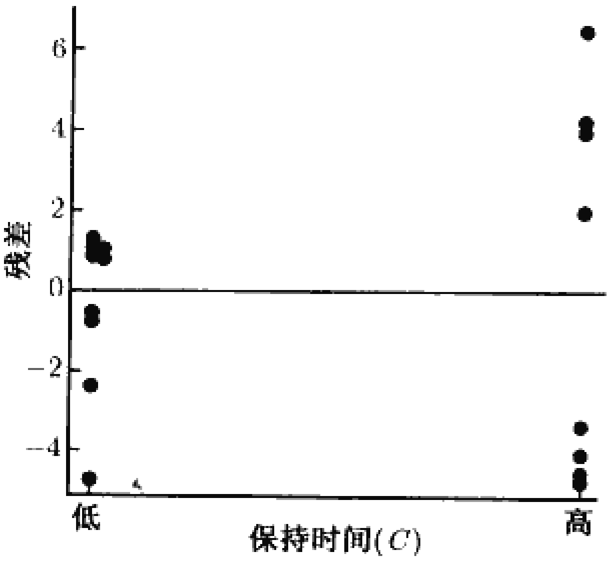


图 8.15 例 8.4 残差与保持时间 (C) 的关系图

更为详细的残差分析如表 8.12 所示. 此表列出了每个因子低水平 (−) 和高水平 (+) 处的残差, 并计算出了每个因子低水平和高水平处的残差的标准差. C 处于低水平时的残差标准差 $[S(C^-)=1.63]$ 明显小于 C 处于高水平时的残差标准差 $[S(C^+)=5.70]$.

表 8.12 下面一行表示统计量

$$F_i^*=\ln\frac{S^2(i^+)}{S^2(i^-)}$$

我们知道, 如果因子 i 处于高 (+) 水平和低 (−) 水平时, 残差的方差相符, 则此比值近似服从均值为零的正态分布, 所以可用它来判断因子 i 处于两个水平时响应变异性的差异. 因为比值 F_C^* 相对较大, 我们的结论是, 从图 8.15 观测到的**分散效应**或**变异性效应**是真实的. 这样一来, 在生产中, 使保持时间处于低水平, 就会减少零件之间收缩率的变异性. 图 8.16 是表 8.12 的 F_i^* 值的正态概率图, 此图也显示出因子 C 有较大的分散效应.

图 8.17 显示了此实验中的数据投影到因子 A,B,C 的立方体上. 收缩率观测值的平均值和极差都在立方体的每个角点上标出. 审查一下这张图就会看到, 进行生产时, 使螺杆转速 (B) 处于低水平, 是减少零件平均收缩率的关键. 实际上, 当 B 处于低水平时, 温度 (A) 和保持时间 (C) 的任一组合都会使零件的平均收缩率取小的数值. 不过, 检查一下立方体每个角点上收缩率的极差, 就会马上看到, 如果在生产时想要保持零件之间收缩率变异性取低值的话, 使保持时间 (C) 处于低水平是唯一合理的选择.

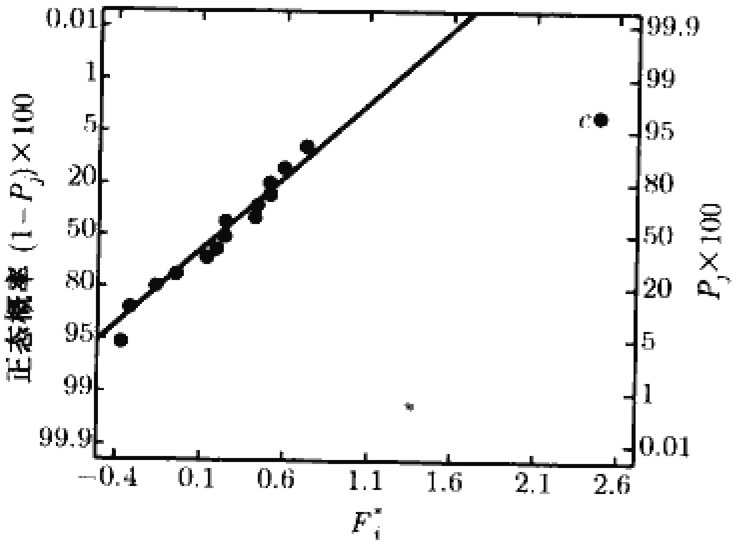


图 8.16 例 8.4 分散效应的正态概率图

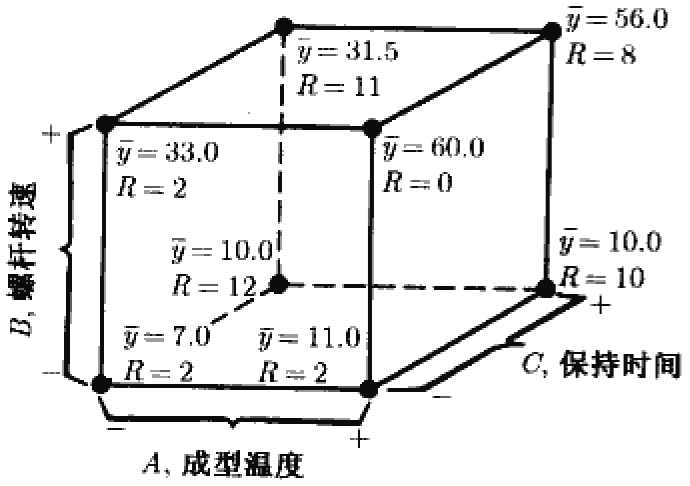


图 8.17 例 8.4 中收缩率关于因子 A,B,C 的平均值和极差

表 8.12　例 8.4 分散效应的计算

试验	A	B	$AB=CE$	C	$AC=BE$	$AE=BC$ $=DF$	E	D	$AD=$ EF	$BD=$ CE	ABD	$BF=$ CD	ACD	F	$AF=DE$	残差
1	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+	−2.50
2	+	−	−	−	−	+	+	−	−	+	+	+	+	−	−	−0.50
3	−	+	−	−	+	−	+	−	+	−	+	+	−	+	−	−0.25
4	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+	2.00
5	−	−	+	+	−	−	+	−	+	+	−	−	+	+	−	−4.50
6	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+	4.50
7	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+	−6.25
8	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	2.00
9	−	−	+	−	+	+	−	+	−	−	+	−	+	+	−	−0.50
10	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+	1.50
11	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+	1.75
12	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−	2.00
13	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	7.50
14	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	−5.50
15	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	4.75
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−6.00
$S(i^+)$	3.80	4.01	4.33	5.70	3.68	3.85	4.17	4.64	3.39	4.01	4.72	4.71	3.50	3.88	4.87	
$S(i^-)$	4.60	4.41	4.10	1.63	4.53	4.33	4.25	3.59	2.75	4.41	3.64	3.65	3.12	4.52	3.40	
F_i^*	−0.38	−0.19	0.11	2.50	−0.42	−0.23	−0.04	0.51	0.42	−0.19	0.52	0.51	0.23	−0.31	0.72	

8.4 一般的 2^{k-p} 分式析因设计

8.4.1 设计的选择

含有 2^{k-p} 个试验的 2^k 分式析因设计叫做 2^k 设计的一个 $1/2^p$ 分式设计, 或简单地说, **2^{k-p} 分式析因设计**. 这些设计需要选取 p 个独立的生成元. 设计的定义关系由原来选好的 p 个生成元以及它们的 2^p-p-1 个广义交互作用所组成. 本节讨论这些设计的结构和分析.

别名结构由各个效应列乘以定义关系求得的. 实际选取生成元时要小心, 不要使感兴趣的效应相互别名. 每一效应有 2^p-1 个别名. 对中等大的 k 值, 通常假定较高阶交互作用 (比方说, 三阶, 四阶或更高) 可被忽略, 这会使别名结构大大简化.

对一个 2^{k-p} 分式析因设计来说, 选好 p 个生成元很重要, 选取的方法要使得能够得到最佳可能的别名关系. 一个合理的标准是, 选取的生成元要使得 2^{k-p} 设计有**最大可能的分辨度**. 为了说明, 考虑表 8.9 中的 2_{IV}^{6-2} 设计, 其中所用的生成元为 $E=ABC$ 和 $F=BCD$, 从而得到分辨度为Ⅳ的设计. 这是分辨度最大的设计. 若选取 $E=ABC$ 和 $F=ABCD$, 完全定义关系是: $I=ABCE=ABCDF=DEF$, 设计的分辨度就会是Ⅲ. 显然, 这是个较差的选择, 因为它牺牲了交互作用的信息.

有时仅靠分辨度还不足以区分设计的优劣. 例如, 考虑表 8.13 中的 3 个 2_{IV}^{7-2} 设计. 这 3 个设计都是分辨度为Ⅳ的, 但在二因子交互作用方面, 它们有很不相同的别名结构 (假定三因子交互作用和更高阶交互作用可以忽略). 显然设计 A 有较多的别名而设计 C 的别名最少, 所以设计 C 是最好的 2_{IV}^{7-2} 选择.

表 8.13 2_{IV}^{7-2} 设计的生成元的 3 种选择

设计 A 的生成元:	设计 B 的生成元:	设计 C 的生成元:
$F=ABC, G=BCD$	$F=ABC, G=ADE$	$F=ABCD, G=ABDE$
$I=ABCF=$	$I=ABCF=$	$I=ABCDF=$
$BCDG=ADFG$	$ADEG=BCDEFG$	$ABDEG=CEFG$
别名 (二因子交互作用)	别名 (二因子交互作用)	别名 (二因子交互作用)
$AB=CF$	$AB=CF$	$CE=FG$
$AC=BF$	$AC=BF$	$CF=EG$
$AD=FG$	$AD=EG$	$CG=EF$
$AG=DF$	$AE=DG$	
$BD=CG$	$AF=BC$	
$BG=CD$	$AG=DE$	
$AF=BC=DG$		

设计 A 中的 3 个字长都是 4, 即, 字长模式为 $\{4,4,4\}$. 设计 B 的是 $\{4,4,6\}$, 而设计 C 的是 $\{4,5,5\}$. 设计 C 只有一个 4 个字母长的字, 而其他设计有两个或三个. 因此, 设计 C 使得定义关系中长度最短的字的个数最少. 我们称这种设计为**最小低阶混杂设计** (minimum aberration design). 在一个分辨度为 R 的设计中, 像差最小的设计保证了设计中与 $R-1$ 阶交互作用有别名的主效应个数最少, 与 $R-2$ 阶交互作用有别名的二因子交互作用个数最少, 等

等. 要了解更多细节, 参见 Fries and Hunter(1980).

对 $k \leqslant 11$ 个因子且最多 $n \leqslant 128$ 个试验, 表 8.14 列出了 2^{k-p} 分式析因设计的选取方法. 此表建议取的生成元可使设计有最大可能的分辨度. 这些也是最小低阶混杂设计.

表 8.14 2^{k-p} 分式析因设计的选取

因子个数, k	分式析因设计	试验个数	设计的生成元
3	2_{III}^{3-1}	4	$C=\pm AB$
4	2_{IV}^{4-1}	8	$D=\pm ABC$
5	2_{V}^{5-1}	16	$E=\pm ABCD$
	2_{III}^{5-2}	8	$D=\pm AB$
			$E=\pm AC$
6	2_{VI}^{6-1}	32	$F=\pm ABCDE$
	2_{IV}^{6-2}	16	$E=\pm ABC$
			$F=\pm BCD$
	2_{III}^{6-3}	8	$D=\pm AB$
			$E=\pm AC$
			$F=\pm BC$
7	2_{VII}^{7-1}	64	$G=\pm ABCDEF$
	2_{IV}^{7-2}	32	$F=\pm ABCD$
			$G=\pm ABDE$
	2_{IV}^{7-3}	16	$E=\pm ABC$
			$F=\pm BCD$
			$G=\pm ACD$
	2_{III}^{7-4}	8	$D=\pm AB$
			$E=\pm AC$
			$F=\pm BC$
			$G=\pm ABC$
8	2_{V}^{8-2}	64	$G=\pm ABCD$
			$H=\pm ABEF$
	2_{IV}^{8-3}	32	$F=\pm ABC$
			$G=\pm ABD$
			$H=\pm BCDE$
	2_{IV}^{8-4}	16	$E=\pm BCD$
			$F=\pm ACD$
			$G=\pm ABC$
			$H=\pm ABD$
9	2_{VI}^{9-2}	128	$H=\pm ACDFG$
			$J=\pm BCEFG$
	2_{IV}^{9-3}	64	$G=\pm ABCD$
			$H=\pm ACEF$
			$J=\pm CDEF$
	2_{IV}^{9-4}	32	$F=\pm BCDE$
			$G=\pm ACDE$
			$H=\pm ABDE$
			$J=\pm ABCE$

因子个数, k	分式析因设计	试验个数	设计的生成元
	2_{III}^{9-5}	16	$E=\pm ABC$
			$F=\pm BCD$
			$G=\pm ACD$
			$H=\pm ABD$
			$J=\pm ABCD$
10	2_{V}^{10-3}	128	$H=\pm ABCG$
			$J=\pm BCDE$
			$K=\pm ACDF$
	2_{IV}^{10-4}	64	$G=\pm BCDF$
			$H=\pm ACDF$
			$J=\pm ABDE$
			$K=\pm ABCE$
	2_{IV}^{10-5}	32	$F=\pm ABCD$
			$G=\pm ABCE$
			$H=\pm ABDE$
			$J=\pm ACDE$
			$K=\pm BCDE$
	2_{III}^{10-6}	16	$E=\pm ABC$
			$F=\pm BCD$
			$G=\pm ACD$
			$H=\pm ABD$
			$J=\pm ABCD$
			$K=\pm AB$
11	2_{IV}^{11-5}	64	$G=\pm CDE$
			$H=\pm ABCD$
			$J=\pm ABF$
			$K=\pm BDEF$
			$L=\pm ADEF$
	2_{IV}^{11-6}	32	$F=\pm ABC$
			$G=\pm BCD$
			$H=\pm CDE$
			$J=\pm ACD$
			$K=\pm ADE$
			$L=\pm BDE$
	2_{III}^{11-7}	16	$E=\pm ABC$
			$F=\pm BCD$
			$G=\pm ACD$
			$H=\pm ABD$
			$J=\pm ABCD$
			$K=\pm AB$

因子个数, k	分式析因设计	试验个数	设计的生成元
			$L=\pm AC$
12	2_{III}^{12-18}	16	$E=\pm ABC$
			$F=\pm ABD$
			$G=\pm ACD$
			$H=\pm BCD$
			$J=\pm ABCD$
			$K=\pm AB$
			$L=\pm AC$
			$M=\pm AD$
13	2_{III}^{13-9}	16	$E=\pm ABC$
			$F=\pm ABD$
			$G=\pm ACD$
			$H=\pm BCD$
			$J=\pm ABCD$
			$K=\pm AB$
			$L=\pm AC$
			$M=\pm AD$
			$N=\pm BC$
14	2_{III}^{14-10}	16	$E=\pm ABC$
			$F=\pm ABD$
			$G=\pm ACD$
			$H=\pm BCD$
			$J=\pm ABCD$
			$K=\pm AB$
			$L=\pm AC$
			$M=\pm AD$
			$N=\pm BC$
			$O=\pm BD$
15	2_{III}^{15-11}	16	$E=\pm ABC$
			$F=\pm ABD$
			$G=\pm ACD$
			$H=\pm BCD$
			$J=\pm ABCD$
			$K=\pm AB$
			$L=\pm AC$
			$M=\pm AD$
			$N=\pm BC$
			$O=\pm BD$
			$P=\pm CD$

当 $n \leqslant 64$ 时, 表 8.14 的所有设计的别名关系列于附录的表 X (a − w). 此表中列出的别名关系着重于主效应、二因子交互作用, 以及三因子交互作用. 对于每一设计给出了完全定义关系. 此附录表使我们非常容易选出一个有足够分辨度的设计, 以确保任意一个可能感兴趣的交互作用能够估计出来.

例 8.5 为了说明表 8.14 的用法, 假定我们有 7 个因子, 并感兴趣于 7 个主效应的估计以及关于二因

子交互作用的某些信息. 假定三因子交互作用和更高阶的交互作用可被忽略. 这些信息提示分辨度为Ⅳ的设计会是合适的.

表 8.14 表明, 有两个可用的分辨度为Ⅳ的分式设计: 有 32 个试验的 2_{IV}^{7-2} 和有 16 个试验的 2_{IV}^{7-3}. 附录的表Ⅹ中有这两个设计的完全别名关系. 有 16 个试验的 2_{IV}^{7-3} 设计的别名列于附录的表Ⅹ (i). 所有 7 个主效应的别名是三因子交互作用. 二因子交互作用的别名还是二因子交互作用, 它们 3 个一组. 因此, 此设计符合我们的目标, 也就是说, 可以估计主效应, 并且给出一些二因子交互作用的信息. 没有必要再进行有 32 个试验的 2_{IV}^{7-2} 设计了. 从附录的表Ⅹ (j) 看到, 此设计可以估计出所有 7 个主效应, 且 21 个二因子交互作用中的 15 个可单独地估计出来 (三因子及更高阶的交互作用被忽略了). 这比所需要的关于交互作用的信息更多. 完整的设计见表 8.15. 构造方法是, 先写出 A, B, C, D 的有 16 次试验的基本设计, 然后加上三列 $E = ABC, F = BCD, G = ACD$. 生成元是 $I = ABCE, I = BCDF, I = ACDG$ (表 8.14). 完全定义关系是 $I = ABCE = BCDF = ADEF = ACDG = BDEG = CEFG = ABFG$.

表 8.15 2_{IV}^{7-3} 分式析因设计

试验	基本设计				$E = ABC$	$F = BCD$	$G = ACD$
	A	B	C	D			
1	−	−	−	−	−	−	−
2	+	−	−	−	+	−	+
3	−	+	−	−	+	+	−
4	+	+	−	−	−	+	+
5	−	−	+	−	+	+	+
6	+	−	+	−	−	+	−
7	−	+	+	−	−	−	+
8	+	+	+	−	+	−	−
9	−	−	−	+	−	+	+
10	+	−	−	+	+	+	−
11	−	+	−	+	+	−	+
12	+	+	−	+	−	−	−
13	−	−	+	+	+	−	−
14	+	−	+	+	−	−	+
15	−	+	+	+	−	+	−
16	+	+	+	+	+	+	+

8.4.2 2^{k-p} 分式析因设计的分析

有许多计算程序可用来分析 2^{k-p} 分式析因设计. 例如, 第 6 章所说的 Design-Expert 就有这种功能.

该设计也可以根据第一原则来分析, 第 i 个效应可用

$$[i] = \frac{2\ (\text{对照}_i)}{N} = \frac{\text{对照}_i}{(N/2)}$$

其中对照 $_i$ 是用第 i 列的加号与减号得到, $N = 2^{k-p}$ 是观测的总个数. 2^{k-p} 分式设计允许估计 $2^{k-p} - 1$ 个效应 (以及它们的别名).

2^{k-p} 分式析因设计的投影

2^{k-p} 设计可压缩为由原来因子的任意 $r \leqslant k - p$ 个子集构成的完全析因设计或分式析因设计. 那些构成分式析因设计的因子子集应是完全定义关系的词中字母的子集. 在筛选实验中, 当我们在进行实验之初就推测原来的多数因子只有较小的效应时, 这类设计特别有用. 这样, 原来

的 2^{k-p} 分式析因设计就可在最感兴趣的因子上投影为完全析因设计. 此类设计所得的结论应认为是暂时的, 须再进一步分析. 有可能得出涉及更高阶交互作用的数据的其他解释.

作为例子, 考虑例 8.5 的 2_{IV}^{7-3} 设计. 这是涉及 7 个因子有 16 个试验的设计. 它可以在原来 7 个因子的任意 4 个因子上投影为完全析因设计, 只要那 4 个因子不是定义关系的词就行. 有 35 个四因子的子集, 其中有 7 个出现在完全定义关系中 (见表 8.15). 这样一来, 还有 28 个四因子子集可构成 2^4 设计. 审查一下表 8.15, 易见一种组合就是 A, B, C, D.

为说明此投影性质的用途, 假定我们进行一个实验来改善球磨机的效率, 7 个因子是: (1) 电动机速度; (2) 增益 (gain); (3) 进料方式; (4) 进料尺寸; (5) 材料类型; (6) 筛网角; (7) 筛网振动水平. 我们相当肯定电动机速度、进料方式、进料尺寸、材料类型会影响效率而且这些因子有交互作用. 其他 3 个因子的作用知道的很少, 但很可能是可忽略的. 一个合理的策略是, 在表 8.15 中分别指定电动机速度、进料方式、进料尺寸和材料类型为列 A, B, C, D. 增益、筛网角、筛网振动水平分别安排在列 E, F, G. 如果我们是对的, 并且 "小变量" E, F, G 可被忽略, 我们就会留下一个关键过程变量中的完全 2^4 设计.

8.4.3　分式析因设计的区组化

有时, 分式析因设计需要做很多试验, 这些试验又不能在相同的条件下做. 此时, 分式析因可能和区组相混. 对于表 8.14 中的许多分式析因设计, 在附录的表 X 中都有推荐的区组化安排. 对这些设计的最小区组大小是 8.

为说明一般方法, 考虑 2_{IV}^{6-2} 分式析因设计, 其定义关系为: $I = ABCE = BCDF = ADEF$, 见表 8.10. 此分式设计含有 16 个处理组合. 假定我们想把设计的试验分为两个区组, 每个区组有 8 种处理组合. 在选取与区组相混的交互作用时, 查一下附录表 X (f) 的别名结构, 注意到仅涉及三因子交互作用的有两个别名组. 表中建议取 ABD (及其别名) 与区组相混. 这就得出如图 8.18 所示的两个区组. 主区组包含的处理组合有偶数个字母与 ABD 相同. 这些也就是使 $L = x_1 + x_2 + x_4 = 0 (\text{mod } 2)$ 的处理组合.

区组 1	区组 2
(1)	*ae*
abf	*acf*
cef	*bef*
abce	*bc*
abef	*df*
bde	*abd*
acd	*cde*
bcdf	*abcdef*

图 8.18　ABD 被混杂的二区组 2_{IV}^{6-2} 设计

例 8.6　一台五轴数控机床用来生产用于喷气涡轮机的叶轮. 叶片的外形是一项重要的质量指标. 特别地, 叶片外形和设计外形的偏差倍受注意. 做个实验来确定哪些加工参数会影响外形偏差. 设计选取的 8 个因子如下:

因　子	低水平 (−)	高水平 (+)
$A = x$ 轴偏移 (0.001 英寸)	0	15
$B = y$ 轴偏移 (0.001 英寸)	0	15
$C = z$ 轴偏移 (0.001 英寸)	0	15
$D =$ 刀具卖主	1	2
$E = a$ 轴偏移 (0.001 度)	0	30
$F =$ 主轴速度 (%)	90	110
$G =$ 夹具高度 (0.001 英寸)	0	15
$H =$ 进料速度 (%)	90	110

在每个零件上选一叶片用来检查. 剖面偏差用坐标测量仪测量. 实际剖面和指定剖面之差的标准差为响应变量.

机床有 4 根主轴. 因为主轴间可能会有差异, 工程师把主轴作为区组来处理.

工程师相信, 三因子交互作用和更高阶的交互作用不太重要, 但他们不愿意忽略二因子交互作用. 查表 8.14, 有两种初看合适的设计: 有 16 个试验的 2_{IV}^{8-4} 设计和有 32 个试验的 2_{IV}^{8-3} 设计. 附录的表 X (1) 表明, 如果用有 16 个试验的设计, 有许多二因子交互作用互为别名. 还有, 这个设计不可能实施在这样 4 个区组中: 使 4 个二因子交互作用不与区组混杂. 因此, 实验者决定用 4 个区组的 2_{IV}^{8-3} 设计. 将一个三因子交互作用别名链与一个二因子交互作用 (EH) 及其三因子交互作用别名与区组相混. EH 交互作用是 a 轴偏移和进料速度之间的交互作用, 工程师认为, 这两个变量之间不会有交互作用 (如果 E 与 H 在这种安排下并不满意, 人们常常安排其他的不会有交互作用的两个因子作为这里的 E 与 H).

表 8.16 列出了此设计以及用标准差 $\times 10^3$ 英寸表示的响应变量. 因为响应变量是标准差, 进行分析时最好使用对数变换. 效应估计显示在表 8.17 中. 图 8.19 是效应估计量的正态概率图. 用 ln(标准差 $\times 10^3$) 作为响应变量. 较大的效应只有 $A = x$ 轴偏移, $B = y$ 轴偏移, 以及别名链 $AD+BG$. AD 是 x 轴偏

表 8.16 例 8.6 的四区组 2^{8-3} 设计

试验	基本设计					$F=ABC$	$G=ABD$	$H=BCDE$	区组	实际试验顺序	标准差 ($\times 10^3$in)
	A	B	C	D	E						
1	−	−	−	−	−	−	−	+	3	18	2.76
2	+	−	−	−	−	+	+	+	2	16	6.18
3	−	+	−	−	−	+	+	−	4	29	2.43
4	+	+	−	−	−	−	−	−	1	4	4.01
5	−	−	+	−	−	+	−	−	1	6	2.48
6	+	−	+	−	−	−	+	−	4	26	5.91
7	−	+	+	−	−	−	+	+	2	14	2.39
8	+	+	+	−	−	+	−	+	3	22	3.35
9	−	−	−	+	−	−	+	−	1	8	4.40
10	+	−	−	+	−	+	−	−	4	32	4.10
11	−	+	−	+	−	+	−	+	2	15	3.22
12	+	+	−	+	−	−	+	+	3	19	3.78
13	−	−	+	+	−	+	+	+	3	24	5.32
14	+	−	+	+	−	−	−	+	2	11	3.87
15	−	+	+	+	−	−	−	−	4	27	3.03
16	+	+	+	+	−	+	+	−	1	3	2.95
17	−	−	−	−	+	−	−	−	2	10	2.64
18	+	−	−	−	+	+	+	−	3	21	5.50
19	−	+	−	−	+	+	+	+	1	7	2.24
20	+	+	−	−	+	−	−	+	4	28	4.28
21	−	−	+	−	+	+	−	+	4	30	2.57
22	+	−	+	−	+	−	+	+	1	2	5.37
23	−	+	+	−	+	−	+	−	3	17	2.11
24	+	+	+	−	+	+	−	−	2	13	4.18
25	−	−	−	+	+	−	+	+	4	25	3.96
26	+	−	−	+	+	+	−	+	1	1	3.27
27	−	+	−	+	+	+	−	−	3	23	3.41
28	+	+	−	+	+	−	+	−	2	12	4.30
29	−	−	+	+	+	+	+	−	2	9	4.44
30	+	−	+	+	+	−	−	−	3	20	3.65
31	−	+	+	+	+	−	−	+	1	5	4.41
32	+	+	+	+	+	+	+	+	4	31	3.40

表 8.17　例 8.6 的效应估计、回归系数以及平方和

变量	名称	−1 水平	+1 水平
A	x 轴偏移	0	15
B	y 轴偏移	0	15
C	z 轴偏移	0	15
D	刀具卖主	1	2
E	a 轴偏移	0	30
F	主轴速度	90	110
G	夹具高度	0	15
H	进料速度	90	110
变量 *	**回归系数**	**效应估计**	**平方和**
总平均	1.280 07		
A	0.145 13	0.290 26	0.674 020
B	−0.100 27	−0.200 54	0.321 729
C	−0.012 88	−0.025 76	0.005 310
D	0.054 07	0.108 13	0.093 540
E	−2.531E-04	−5.063E-04	2.050E-06
F	−0.019 36	−0.038 71	0.011 988
G	0.058 04	0.116 08	0.107 799
H	0.007 08	0.014 17	0.001 606
$AB+CF+DG$	−0.002 94	0.005 88	2.767E-04
$AC+BF$	−0.031 03	−0.062 06	0.030 815
$AD+BG$	−0.187 06	−0.374 12	1.119 705
AE	0.004 02	0.008 04	5.170E-04
$AF+BC$	−0.022 51	−0.045 02	0.016 214
$AG+BD$	0.026 44	0.052 88	0.022 370
AH	−0.025 21	−0.050 42	0.020 339
BE	0.049 25	0.098 51	0.077 627
BH	0.006 54	0.013 09	0.001 371
$CD+FG$	0.017 26	0.034 52	0.009 535
CE	0.019 91	0.039 82	0.012 685
$CG+DF$	−0.007 33	−0.014 67	0.001 721
CH	0.030 40	0.060 80	0.029 568
DE	0.008 54	0.017 08	0.002 334
DH	0.007 84	0.015 69	0.001 969
EF	−0.009 04	−0.018 08	0.002 616
EG	−0.026 85	−0.053 71	0.023 078
EH	−0.017 67	−0.035 34	0.009 993
FH	−0.014 04	−0.028 08	0.006 308
GH	0.002 45	0.004 89	1.914E-04
ABE	0.016 65	0.033 31	0.008 874
ABH	−0.006 31	−0.012 61	0.001 273
ACD	−0.027 17	−0.054 33	0.023 617

* 只有主效应与二因子交互作用.

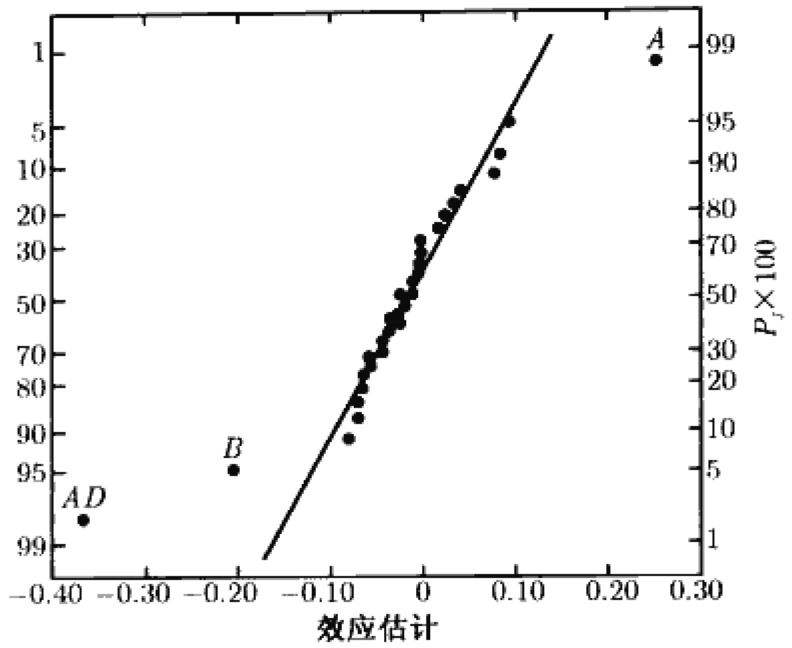


图 8.19 例 8.6 效应估计的正态概率图

移–刀具卖主的交互作用, BG 是 y 轴偏移–夹具高度的交互作用, 因为两个交互作用是混杂的, 不可能根据当前实验数据来区分它们. 又因为两个交互作用各自都包含一个大的主效应, 也难以用任何 "显然的" 简单逻辑来处理这种情况. 若某种工程知识或工序知识对此有帮助, 也许可以在这两个交互作用之间进行选择; 否则, 需要更多的数据来区分这两个效应 (给分式析因设计增加试验来分离交互作用别名的问题将在 8.5 节与 8.6 节中讨论).

假定工序知识指出合适的交互作用可能是 AD. 表 8.18 是这一模型用因子 A, B, D, AD (包含因子 D 是为了遵循分层原则) 的方差分析结果. 区组效应显得较小, 因此可认为机床主轴没有大的差异.

表 8.18 例 8.6 的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
A	0.674 0	1	0.674 0	39.42	< 0.000 1
B	0.321 7	1	0.321 7	18.81	0.000 2
D	0.093 5	1	0.093 5	5.47	0.028 0
AD	1.119 7	1	1.119 7	65.48	< 0.000 1
区组	0.020 1	3	0.006 7		
误差	0.409 9	24	0.017 1		
总和	2.638 9	31			

图 8.20 是此实验中残差的正态概率图. 此图比正态尾部稍重, 所以, 可能要考虑其他变换. 图 8.21 是

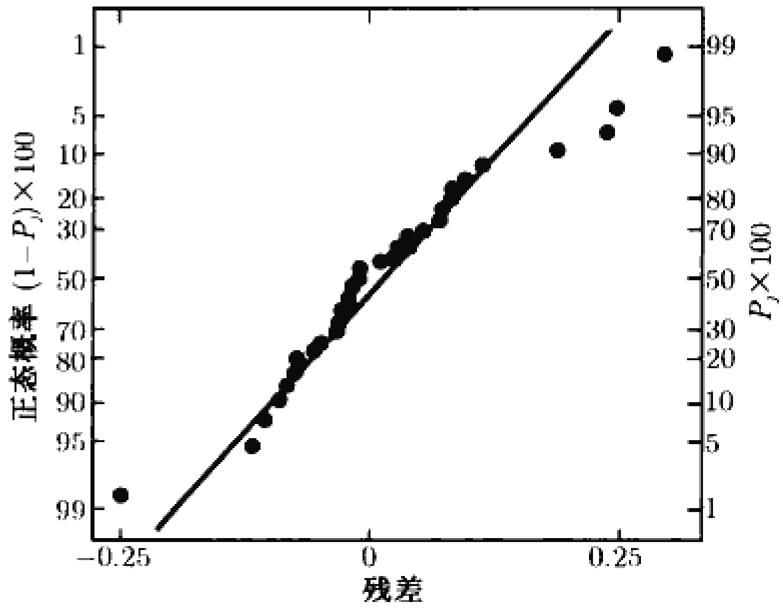


图 8.20 例 8.6 残差的正态概率图

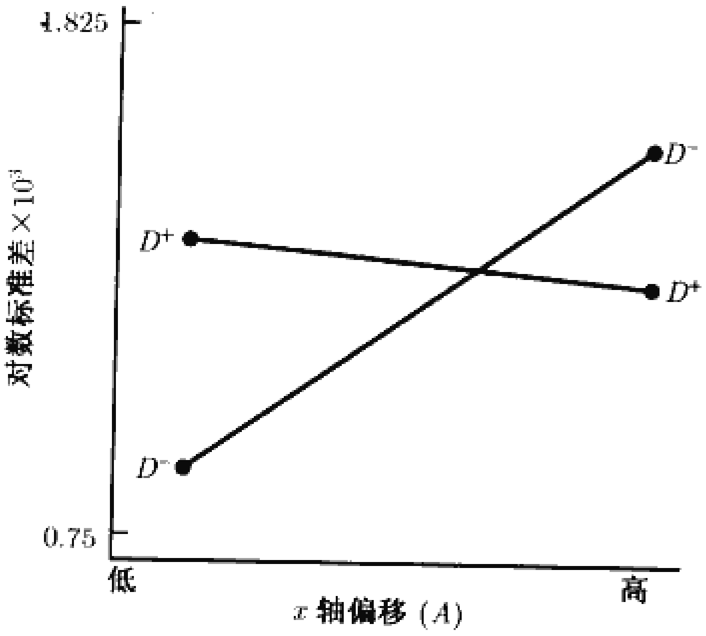


图 8.21 例 8.6 中 AD 交互作用的图形

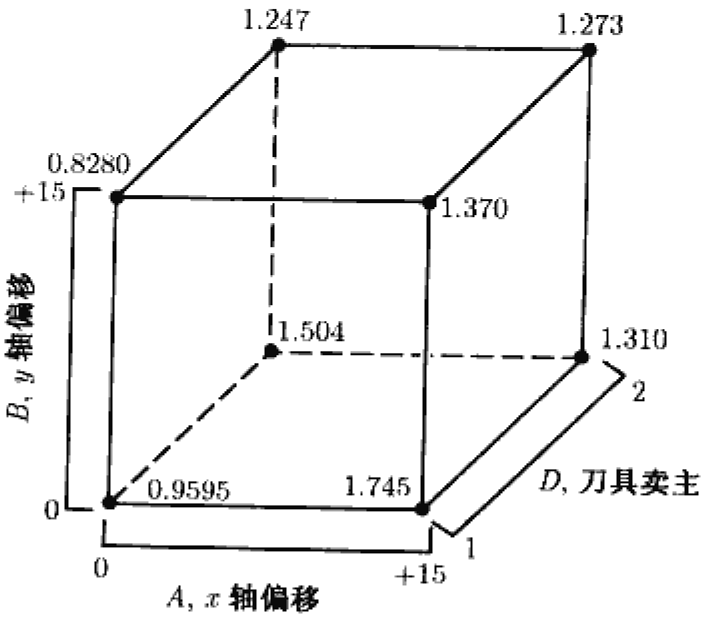


图 8.22 例 8.6 中的 2_{IV}^{8-3} 设计在因子 A, B, D 的有 4 次重复的 2^3 设计上的投影

AD 交互作用图. 刀具卖主 (D) 和 x 轴的偏移 (A) 量深刻地影响到叶片剖面与指定剖面之间的变异性. 做实验时, 令 A 处于低水平 (零偏移) 并从卖主 1 处购买刀具, 将会得到最好的结果. 图 8.22 是此 2_{IV}^{8-3} 设计在因子 A, B, D 的有 4 次重复的 2^3 设计中的投影. 操作条件的最好组合是 A 处于低水平 (0 偏移)、B 处于高水平 (0.015 英寸偏移)、D 处于低水平 (卖主 1 的刀具).

8.5 分辨度为III的设计

8.5.1 分辨度为III的设计的构造

如早先指出的, 序贯地使用分式析因设计非常有用, 通常能够大大节省实验费用和提高实验效率. 分式析因设计的这种应用在纯粹的因子筛选场合经常出现, 即, 有相对较多因子但其中只有很少一些可能是重要的. 分辨度为III的设计在这些场合中非常有用.

对只有 N 个试验和最多可有 $k = N - 1$ 个因子的情况, 有可能构造分辨度为III的设计, 其中 N 是 4 的倍数. 这些设计常用在工业实验中. N 是 2 的幂的设计, 可以用本章前面介绍的方法构造出来, 不过, 这类方法是最先介绍的. 特别重要的设计是最多为 3 个因子但只要 4 个试验的设计, 最多为 7 个因子但只要 8 个试验的设计, 最多为 15 个因子但只要 16 个试验的设计. 当 $k = N - 1$ 时, 分式析因设计叫做**饱和的**.

用于分析至多 3 个因子 4 个试验的设计是原先 8.2 节中介绍过的 2_{III}^{3-1} 设计. 另一个十分有用的饱和分式析因设计是研究有 7 个因子 8 个试验的设计, 即 2_{III}^{7-4} 设计. 此设计是 2^7 的 1/16 分式. 其构造方法是, 作为基本设计, 先写 A, B, C 完全 2^3 设计的加号水平和减号水平, 然后加进 4 个附加因子的水平, 此 4 个因子是原来 3 个因子的交互作用, 具体如下: $D = AB, E = AC, F = BC, G = ABC$. 于是, 这一设计的生成元是 $I = ABD, I = ACE, I = BCF, I = ABCG$. 此设计见表 8.19.

此设计的完全定义关系, 是通过将 4 个生成元 ABD、ACE、BCF、$ABCG$ 两两相乘、三三相乘、四四相乘而求得, 具体是

$$
\begin{aligned}
I &= ABD = ACE = BCF = ABCG = BCDE = ACDF = CDG \\
&= ABEF = BEG = AFG = DEF = ADEG = CEFG = BDFG = ABCDEFG
\end{aligned}
$$

表 8.19 生成元为 $I = ABD$、$I = ACE$、$I = BCF$、$I = ABCG$ 的 2_{III}^{7-4} 设计

试验	基本设计			$D = AB$	$E = AC$	$F = BC$	$G = ABC$	
	A	B	C					
1	−	−	−	+	+	+	−	def
2	+	−	−	−	−	+	+	afg
3	−	+	−	−	+	−	+	beg
4	+	+	−	+	−	−	−	abd
5	−	−	+	+	−	−	+	cdg
6	+	−	+	−	+	−	−	ace
7	−	+	+	−	−	+	−	bcf
8	+	+	+	+	+	+	+	$abcdefg$

为找出任一效应的别名, 只需将定义关系中的每个词乘该效应即可. 例如, B 的别名是

$$
\begin{aligned}
B &= AD = ABCE = CF = ACG = CDE = ABCDF = BCDG = AEF = EG \\
&= ABFG = BDEF = ABDEG = BCEFG = DFG = ACDEFG
\end{aligned}
$$

此设计是 1/16 分式设计, 因为所选的生成元取正号, 所以是主分式. 它又是分辨度为III的, 因为它的定义关系中任一词的字母最小个数是 3. 此族 16 个不同的 2_{III}^{7-4} 设计中的任一设计, 都可以用生成元 $I = \pm ABD$、$I = \pm ACE$、$I = \pm BCF$、$I = \pm ABCG$ 中符号的 16 种可能的排列之一构造出来.

该设计的 7 个自由度可用来估计 7 个主效应. 这些效应中的每一个都有 15 个别名. 不过, 如果假定三因子交互作用及更高阶的交互作用可被忽略, 则别名结构将大大简化. 用此假定, 与此设计的 7 个主效应有关的每个线性组合, 实际上估计了主效应和 3 个二因子交互作用:

$$
\begin{aligned}
&[A] \to A + BD + CE + FG \qquad &[E] \to E + AC + BG + DF \\
&[B] \to B + AD + CF + EG \qquad &[F] \to F + BC + AG + DE \\
&[C] \to C + AE + BF + DG \qquad &[G] \to G + CD + BE + AF \\
&[D] \to D + AB + CG + EF &
\end{aligned}
\tag{8.1}
$$

这些别名可在附录的表 X (h) 中查得, 略去了三因子交互作用和更高阶的交互作用.

表 8.19 的饱和 2_{III}^{7-4} 设计可用来求出分辨度为III的少于 7 个因子的 8 个试验的设计. 例如, 要生成一个 6 个因子 8 个试验的设计, 只要简单地删去表 8.19 中的任一列就行, 例如, 列 G. 所得的设计见表 8.20.

容易证明, 此设计也是分辨度为III的. 实际上, 它是一个 2_{III}^{6-3} 设计或 2^6 设计的 1/8 分式设计. 2_{III}^{6-3} 设计的定义关系等于从原来的 2_{III}^{7-4} 设计的定义关系中删去那些含有字母 G 的词后留下来的词. 这样一来, 新设计的定义关系是

$$
I = ABD = ACE = BCF = BCDE = ACDF = ABEF = DEF
$$

一般说来. 删去 d 个因子后所得的新设计, 其新的定义关系只要从原来的定义关系中删去含有被删字母的那些词即可求得. 在用这一方法构造设计时, 要小心地操作以求得最好可能的排列方

式. 如果从表 8.19 中删去列 B, D, F, G 时, 则会得出一个三因子 8 个试验的设计. 不过, 该处理组合对应于 2^{3-1} 设计的两次重复. 实验者可能更愿意做 A, C, E 的完全 2^3 设计.

表 8.20 生成元为 $I = ABD$、$I = ACE$、$I = BCF$ 的 2_{III}^{6-3} 设计

试验	基本设计			$D = AB$	$E = AC$	$F = BC$	
	A	B	C				
1	−	−	−	+	+	+	def
2	+	−	−	−	−	+	af
3	−	+	−	−	+	−	be
4	+	+	−	+	−	−	abd
5	−	−	+	+	−	−	cd
6	+	−	+	−	+	−	ace
7	−	+	+	−	−	+	bcf
8	+	+	+	+	+	+	$abcdef$

也可求得至多 15 个因子的 16 个试验的分辨度为III的设计. 先生成饱和的 2_{III}^{15-11} 设计, 方法是, 先写出有 16 个处理组合的 A, B, C, D 的 2^4 设计, 然后加进 11 个新的因子, 它们是原来的 4 个因子的二因子交互作用、三因子交互作用、四因子交互作用. 在此设计中, 15 个主效应的每个的别名是 7 个二因子交互作用. 用同样的方法求得 2_{III}^{31-26} 设计, 它允许在 32 次试验中最多研究 31 个因子.

8.5.2 折叠分辨度为III的分式设计以分离别名效应

利用把有某些符号变换后的分式析因设计组合起来的方法, 可以系统地把所感兴趣的效应隔离开来. 这种序贯实验称为原先设计的**折叠** (fold over). 在原来的分式设计的别名结构中, 对某一恰当的因子改变它的符号, 就可以求得有一个或多个因子为反号的分式设计的别名结构.

考虑表 8.19 的 2_{III}^{7-4} 设计. 在这一主分式设计中, 使因子 D 的列取反号, 就得出第二个分式设计. 也就是说, 第二个分式结构的 D 列是

$$- + + - - + + -$$

从第一个分式设计, 效应可用 (8.1) 式来估计. 假定三因子交互作用与更高阶交互作用不显著, 则由第二个分式设计得

$$\begin{aligned}
&[A]' \to A - BD + CE + FG && [-D]' \to -D + AB + CG + EF \\
&[B]' \to B - AD + CF + EG && [E]' \to E + AC + BG - DF \\
&[C]' \to C + AE + BF - DG && [F]' \to F + BC + AG - DE \\
&[D]' \to D - AB - CG - EF && [G]' \to G - CD + BE + AF
\end{aligned} \tag{8.2}$$

现在, 由两个效应的线性组合 $\frac{1}{2}([i] + [i]')$ 和 $\frac{1}{2}([i] - [i]')$, 可以求得

i	由 $\frac{1}{2}([i] + [i]')$	由 $\frac{1}{2}([i] - [i]')$
A	$A + CE + FG$	BD
B	$B + CF + EG$	AD
C	$C + AE + BF$	DG
D	D	$AB + CG + EF$
E	$E + AC + BG$	DF
F	$F + BC + AG$	DE
G	$G + BE + AF$	CD

这样一来, 我们就把主效应 D 以及它的所有二因子交互作用分离出来. 一般说来, 如果对一个分辨度为III的分式析因设计或更高级的分式设计, 加进与它的单因子符号相反的设计, 则组合的设计可求得那个因子的主效应及其二因子交互作用的估计量. 这种设计有时被称为**单因子折叠设计**.

现在, 假定我们对一个分辨度为III的分式析因设计加进第二个分式设计, 此设计的所有因子都取原设计的相反符号. 这种折叠 (有时称为**完全折叠, 或反射**) 会打破主效应和二因子交互作用之间的别名联系. 即, 我们可以用**组合的设计**来估计与任何二因子交互作用脱离开来的所有主效应. 下面的例子说明这一完全折叠方法.

例 8.7 一位人类行为分析师进行一项实验来研究眼睛聚焦时间, 并设计了一种仪器. 在测试时, 有几个因子可以控制. 他原来认为重要的因子是: 分辨能力或视觉敏锐度 (A), 目标至眼睛的距离 (B), 目标形状 (C), 照明度水平 (D), 目标大小 (E), 目标密度 (F) 以及试验对象 (G). 每个因子考虑两个水平. 他推测这 7 个因子中仅有少数几个最为重要, 并且因子之间的高阶交互作用可被忽略. 根据这一假定, 分析师决定进行筛选实验以便识别出最重要的因子, 然后进一步集中研究这些因子. 为了筛选这 7 个因子. 他用如表 8.19 所示的 2_{III}^{7-4} 设计, 依随机顺序对处理组合进行试验, 得出以微秒为单位的聚焦时间如表 8.21 所示.

表 8.21 眼睛聚焦时间实验的 2_{III}^{7-4} 设计

试验	基本设计			$D=AB$	$E=AC$	$F=BC$	$G=ABC$		时间
	A	B	C						
1	−	−	−	+	+	+	−	def	85.5
2	+	−	−	−	−	+	+	afg	75.1
3	−	+	−	−	+	−	+	beg	93.2
4	+	+	−	+	−	−	−	abd	145.4
5	−	−	+	+	−	−	+	cdg	83.7
6	+	−	+	−	+	−	−	ace	77.6
7	−	+	+	−	−	+	−	bcf	95.0
8	+	+	+	+	+	+	+	$abcdefg$	141.8

用这些数据可估计 7 个主效应和它们的别名. 由 (8.1) 式可知这些效应及其别名为:

$$[A]=20.63 \to A+BD+CE+FG \qquad [E]=-0.28 \to E+AC+BG+DF$$
$$[B]=38.38 \to B+AD+CF+EG \qquad [F]=-0.63 \to F+BC+AG+DE$$
$$[C]=-0.28 \to C+AE+BF+DG \qquad [G]=-2.43 \to G+CD+BE+AF$$
$$[D]=28.88 \to D+AB+CG+EF$$

例如, A 的主效应和它的别名是:

$$[A]=\frac{1}{4}(-85.5+75.1-93.2+145.4-83.7+77.6-95.0+141.8)=20.63$$

3 个最大的效应是 $[A],[B],[D]$. 对这些数据最简单的解释是主效应 A,B,D 都是显著的. 不过, 这一解释并不是唯一的. 因为, 从逻辑上来说, 结论也可以是 A,B,AB 交互作用, 或者也许是 B,D,BD 交互作用, 或者也许是 A,D,AD 交互作用是真正的效应.

ABD 是此设计的定义关系中的一个词. 因此, 这一 2_{III}^{7-4} 设计不能投影到 ABD 的完全 2^3 析因设计中去, 而是投影到有两次重复的 2^{3-1} 设计中去, 如图 8.23 所示的那样. 因为 2^{3-1} 分式设计是分辨度为III的设计, A 的别名是 BD, B 的别名是 AD, D 的别名是 AB, 所以, 交互作用不能与主效应分离. 此处, 实验者并不走运. 如果他分派照明度水平给 C 而不是 D 的话, 则设计可投影到完全 2^3 设计中, 解释就可以更简单些.

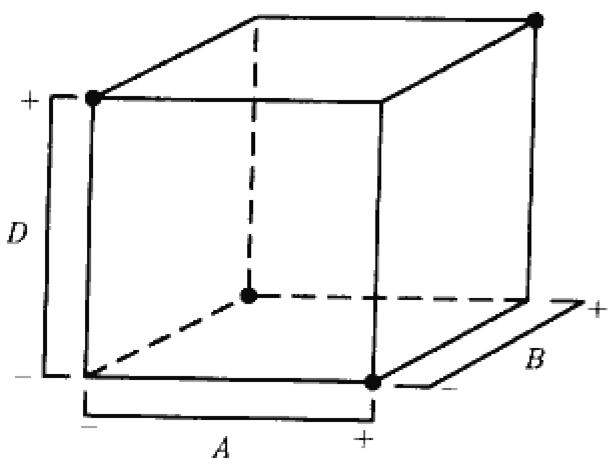


图 8.23 2_{III}^{7-4} 设计投影到 A,B,D 的 2_{III}^{3-1} 设计两次重复中

为把主效应和二因子交互作用分离开来，可以用完全折叠的方法，做第二个所有符号都相反的分式设计．这一折叠设计及其响应变量观测值如表 8.22 所示．当用这个方法折叠一分辨度为III的设计时，我们 (依效应的意义) 改变了有奇数个字母的生成元的符号．用此分式设计估计的效应是

$$[A]' = -17.68 \to A - BD - CE - FG$$
$$[B]' = 37.73 \to B - AD - CF - EG$$
$$[C]' = -3.33 \to C - AE - BF - DG$$
$$[D]' = 29.88 \to D - AB - CG - EF$$
$$[E]' = 0.53 \to E - AC - BG - DF$$
$$[F]' = 1.63 \to F - BC - AG - DE$$
$$[G]' = 2.68 \to G - CD - BE - AF$$

表 8.22 眼睛聚焦实验的一个 2_{III}^{7-4} 折叠设计

试验	基本设计			$D=-AB$	$E=-AC$	$F=-BC$	$G=ABC$		时间
	A	B	C						
1	+	+	+	−	−	−	+	*abcg*	91.3
2	−	+	+	+	+	−	−	*bcde*	136.7
3	+	−	+	+	−	+	−	*acdf*	82.4
4	−	−	+	−	+	+	+	*cefg*	73.4
5	+	+	−	−	+	+	−	*abef*	94.1
6	−	+	−	+	−	+	+	*bdfg*	143.8
7	+	−	−	+	+	−	+	*adeg*	87.3
8	−	−	−	−	−	−	−	(1)	71.9

将第二个分式设计和原来的分式设计组合起来，就求得下面的效应估计量：

i	由 $\frac{1}{2}([i]+[i]')$	由 $\frac{1}{2}([i]-[i]')$
A	$A=1.48$	$BD+CE+FG=19.15$
B	$B=38.05$	$AD+CF+EG=0.33$
C	$C=-1.80$	$AE+BF+DG=1.53$
D	$D=29.38$	$AB+CG+EF=-0.50$
E	$E=0.13$	$AC+BG+DF=-0.40$
F	$F=0.50$	$BC+AG+DE=-1.53$
G	$G=0.13$	$CD+BE+AF=-2.55$

两个最大的效应是 B 和 D．而第三大效应是 $BD+CE+FG$，因此把它归之于 BD 交互作用看来是合理的．在随后的实验中，实验者用距离 (B) 和照明水平 (D) 这两个因子，而将其他因子 A,C,E,F 固定在标准状态，所得的结果证实了上面的分析．为不忽视潜在的对象效应，他决定在新的一轮实验中把试验对象作为区组．因为需要几个不同的对象来完成这个实验．

1. 折叠设计的定义关系

如例 8.7 所示，通过折叠将分式析因设计组合的方法是十分有用的．通常需要了解组合设计的定义关系．这很容易确定．各个分式设计都有 $L+U$ 个词用作生成元：L 个符号相同的词，U 个符号不同的词．组合起来的设计将有 $L+U-1$ 个词作为生成元．其中 L 个词符号相同，而

另外的 $U-1$ 个词的构成方法是, 取符号不相同的那些词中的偶数个相乘所得出的独立词 (**偶数个相乘**的意思是一次取两个词相乘, 一次取 4 个词相乘, 等等).

为说明这一方法, 考虑例 8.7 的设计. 第一个分式设计的生成元是

$$I = ABD, I = ACE, I = BCF, I = ABCG$$

第二个分式设计的生成元是

$$I = -ABD, I = -ACE, I = -BCF, I = ABCG$$

注意, 在第二个分式设计中, 已经把奇数个字母的生成元的符号转换了. 现在, $L+U=1+3=4$. 组合的设计将有 $I = ABCG$ (相同符号的词) 作为一个生成元, 以及独立的由不同符号的偶数个词相乘得的两个词. 例如, 取 $I = ABD$ 和 $I = ACE$, 则 $I = (ABD)(ACE) = BCDE$ 是组合的设计的一个生成元. 还有, 取 $I = ABD$ 和 $I = BCF$, 则 $I = (ABD)(BCF) = ACDF$ 也是组合的设计的一个生成元. 组合的设计的完全定义关系是

$$I = ABCG = BCDE = ACDF = ADEG = BDFG = ABEF = CEFG$$

2. 折叠设计中的区组化

通常折叠设计在两个不同的时间段实施. 在最初的分式设计之后, 进行数据分析与安排折叠设计一般会花费一些时间. 然后才实施第二组试验, 通常在另一天, 或另一班次, 或用不同的操作人员, 或使用不同来源的材料. 因此需要用**区组化**消除两个时段间潜在的讨厌效应. 幸运的是, 组合实验中的区组化是很容易完成的.

以例 8.7 的折叠设计来说明. 在表 8.21 所列的最初 8 次试验中, 生成元是 $D=AB, E=AC, F=BC, G=ABC$. 在表 8.22 折叠设计的试验安排中, 有 3 个生成元的符号改变, 使得 $D=-AB$, $E=-AC$, $F=-BC$. 这样, 最初的 8 次试验中, 效应 ABD、ACE、BCF 的符号都是正的; 而在后面的 8 次试验中, 效应 ABD、ACE、BCF 的符号都是负的. 因此, 这些效应与区组混杂. 实际上, 这是一个单自由度的别名链与区组混杂 (因为只有两个区组, 所以必须有一个自由度用于区组), 且别名链中的效应, 可以根据设计的定义关系, 通过乘以 ABD、ACE、BCF 中的任意一个来得到. 求得:

$$ABD = CDG = ACE = BCF = BEG = AFG = DEF = ABCDEFG$$

是所有与区组混杂的效应. 一般而言, 一个完全的折叠实验总会形成两个区组, 若有效应 (以及它们的别名) 在一个区组中符号为正且在另一个区组中符号为负, 则它们会与区组混杂. 这些效应总可以由那些改变符号而形成折叠的生成元来确定.

8.5.3 Plackett-Burman 设计

Plackett and Burman (1946) 提出的这些设计, 是为在 N 次试验中研究 $k = N-1$ 个变量的二水平分式析因设计, 其中 N 是 4 的倍数. 如果 N 是 2 的幂, 这些设计就是本节前面出现过的设计. 但是, 当 $N = 12$、20、24、28 和 36 时, 就会对 Plackett-Burman 设计感兴趣了. 因为这些设计不能表示为立方体, 所以它们有时被称为**非几何设计** (nongeometric design).

表 8.23 上半部分列出了各加减符号的行, 它们用于构造 $N=12$、20、24 和 36 时的 Plackett-Burman 设计, 而表的下半部分分块列出的加减符号是为 $N=28$ 时构造设计用的. $N=12$、20、24 和 36 的设计的求法, 是将表 8.23 中相应的一行写为一列 (或行), 然后生成第二列 (或行), 方法是将第一列 (行) 中的元素每个下移 (或右移) 一个位置, 将最后一个元素移至第一个位置. 第三列 (或行) 用同样的方法由第二列 (或行) 求得, 直到生成 k 列 (或行) 为止. 然后加进一行减号, 完成这一设计. 对 $N=28$, 3 个区组 X、Y 和 Z 写为如下次序:

$$\begin{matrix} X & Y & Z \\ Z & X & Y \\ Y & Z & X \end{matrix}$$

并在这 27 行下加进一行减号. $N=12$ 个试验和 $k=11$ 个因子的设计见表 8.24.

表 8.23　Plackett-Burman 设计的加减符号

$k=11, N=12$ + + − + + + − − − + −

$k=19, N=20$ + + − − + + + + − + − + − − − − + + −

$k=23, N=24$ + + + + + − + − + + − − + + − − + − + − − − −

$k=35, N=36$ − + − + + + − − − + + + + + − + + + − − + − − − − + − + − + + − − + −

$k=27, N=28$

+ − + + + + − − −	− + − − − + − − +	+ + − + − + + − +
+ + − + + + − − −	− − + + − − + − −	− + + + + − + + −
− + + + + + − − −	+ − − − + − − + −	+ − + − + + − + +
− − − + − + + + +	− − + − + − − − +	+ − + + + − + − +
− − − + + − + + +	+ − − − − + + − −	+ + − − + + + + −
− − − − + + + + +	− + − + − − − + −	− + + + − + − + +
+ + + − − − + − +	− − + − − + − + −	+ − + + − + + + −
+ + + − − − + + −	+ − − + − − − − +	+ + − + + − − + +
+ + + − − − − + +	− + − − + − + − −	− + + − + + + − +

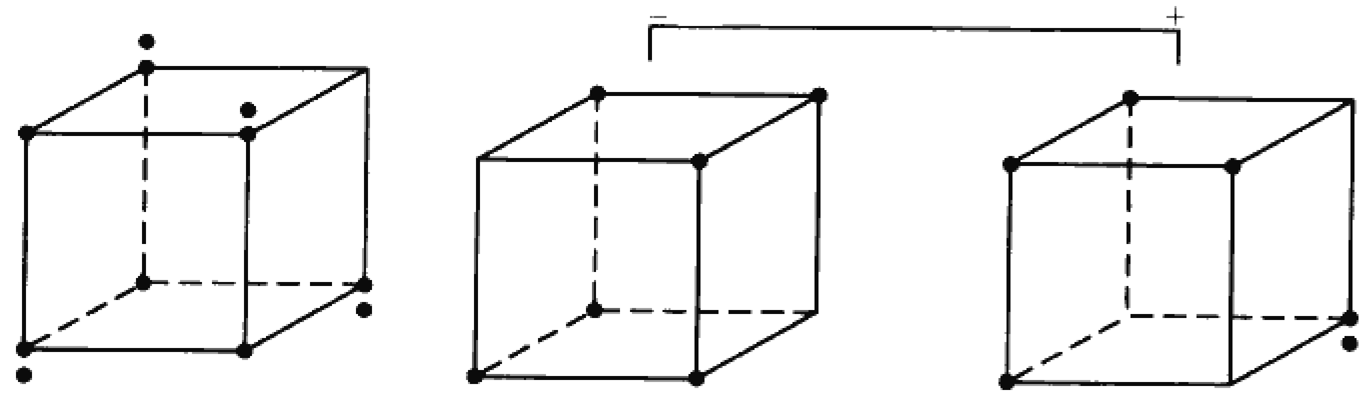
表 8.24　关于 $N=12, k=11$ 的 Plackett-Burman 设计

试验	A	B	C	D	E	F	G	H	I	J	K
1	+	−	+	−	−	−	+	+	+	−	+
2	+	+	−	+	−	−	−	+	+	+	−
3	−	+	+	−	+	−	−	−	+	+	+
4	+	−	+	+	−	+	−	−	−	+	+
5	+	+	−	+	+	−	+	−	−	−	+
6	+	+	+	−	+	+	−	+	−	−	−
7	−	+	+	+	−	+	+	−	+	−	−
8	−	−	+	+	+	−	+	+	−	+	−
9	−	−	−	+	+	+	−	+	+	−	+
10	+	−	−	−	+	+	+	−	+	+	−
11	−	+	−	−	−	+	+	+	−	+	+
12	−	−	−	−	−	−	−	−	−	−	−

$N=12$、20、24、28 和 36 的 Plackett-Burman 设计的别名结构十分凌乱. 例如, 在 12 个试验的设计中, 每一个主效应部分地与每个不涉及它本身的二因子交互作用有别名关系. 例如,

AB 交互作用与 9 个主效应 $C, D, \cdots, K$ 有**部分别名**关系. 而且, 每个主效应与 45 个二因子交互作用有**部分别名**关系. 在较大的设计中, 情况更为复杂. 我们建议实验者, 用这些设计时要非常小心.

非几何 Plackett-Burman 设计的投影性质是有趣的, 在有些时候也很有用. 例如, 考虑表 8.24 的 12 个试验的设计. 此设计会投影为原来 11 个因子中任意两个因子的有 3 次重复的完全 2^2 设计. 但是, 在三因子中, 投影设计是一完全 2^3 析因设计加一个 2_{III}^{3-1} 分式析因设计 (见图 8.24a). 从而, 分辨度为III的 Plackett-Burman 设计有**投影性 3**, 这是指它能拆解成任意三因子的完全析因设计. 而 2_{III}^{k-p} 设计则只有投影性 2. 4 维投影见图 8.24b. 这些四因子投影都不是平衡设计.



(a) 投影为三因子　　(b) 投影为四因子

图 8.24　12 个试验的 Plackett-Burman 设计投影为三因子设计和四因子设计

例 8.8 我们用 11 个变量 12 个试验的设计以及一组模拟数据, 来说明与 Plackett-Burman 设计有关的一些难点. 假定过程有 3 个显著的主效应 (A, B, D) 和两个显著的二因子交互作用 (AB 和 AD). 假定模型为

$$y = 200 + 8x_1 + 10x_2 + 12x_4 - 12x_1x_2 + 9x_1x_4 + \varepsilon$$

其中每个 x_i 都是定义在 -1 到 1 区间上的规范变量, ε 是 NID(0,9) 的随机误差项. 这样, $k = 11$ 个因子中有 3 个是大的, 二因子交互作用中有两个是大的, 这种情况是合理的.

表 8.25 给出 12 次的 Plackett-Burman 设计和模拟响应值. 此设计看起来与表 8.24 中的 12 次设计不同, 是因为它用表 8.23 中 $k = 11, N = 12$ 的行的符号作为表 8.25 的一行. 表 8.26 列出了效应的估计.

表 8.25　例 8.8 的 Plackett-Burman 设计

试验	A	B	C	D	E	F	G	H	I	J	K	响应值
1	+	+	−	+	+	+	−	−	−	+	−	231
2	−	+	+	−	+	+	+	−	−	−	+	207
3	+	−	+	+	−	+	+	+	−	−	−	230
4	−	+	−	+	+	−	+	+	+	−	−	217
5	−	−	+	−	+	+	−	+	+	+	−	175
6	−	−	−	+	−	+	+	−	+	+	+	176
7	+	−	−	−	+	−	+	+	−	+	+	183
8	+	+	−	−	−	+	−	+	+	−	+	185
9	+	+	+	−	−	−	+	−	+	+	−	181
10	−	+	+	+	−	−	−	+	−	+	+	220
11	+	−	+	+	+	−	−	−	+	−	+	229
12	−	−	−	−	−	−	−	−	−	−	−	168

表 8.26　例 8.8 的效应估计、回归系数以及平方和

变量 *	回归系数	效应估计	平方和	变量 *	回归系数	效应估计	平方和
总平均值	200.167			F	0.500	1.000	3.000
A	6.333	12.667	481.333	G	−1.167	−2.333	16.333
B	6.667	13.333	533.333	H	1.500	3.000	27.000
C	6.833	12.667	560.333	J	−6.333	−12.667	481.333
D	17.000	34.000	3 468.000	K	−5.833	−11.667	408.333
E	6.833	13.667	560.333	L	−0.167	−0.333	0.333

* 每个主效应与 45 个二因子交互作用有部分别名关系.

可以看到有 7 个大的效应; A, B, C, D, E, J, K (当然, 还要包括它们的别名). 但不能立即看出其中哪些效应可能是交互作用. 其中有些混杂可以通过折叠设计来区分. 这种做法可以区分主效应, 但一般实验者对交互作用效应仍无法确定.

上面例子所指出的 Plackett-Burman 设计在解释中的困难, 在实际中经常出现. 带有 16 次试验的几何的 2_{III}^{11-7} 设计与一个 12 次试验的可能需要折叠 (这样就要 24 次) 的 Plackett-Burman 设计相比较, 几何设计也许是较好的选择. 更多内容见 Montgomery, Brror and Stanley (1997-1998). 在某些情况下, 可以用回归建模技术处理非几何的 Plackett-Burman 设计中的混杂. Hamada and Wu(1992) 对此有过讨论.

8.6　分辨度为Ⅳ和Ⅴ的设计

8.6.1　分辨度为Ⅳ的设计

当主效应和二因子交互作用是纯净的, 且某些二因子交互作用互为别名时, 2^{k-p} 分式析因设计的分辨度为Ⅳ. 于是, 当删掉三因子交互作用和更高阶的交互作用时, 在 2_{IV}^{k-p} 设计中就可以直接估计主效应. 表 8.10 的 2_{IV}^{6-2} 设计就是一例. 还有, 例 8.7 中两个组合的 2_{IV}^{7-4} 分式设计生成一个 2_{IV}^{7-3} 设计. 分辨度为Ⅳ的设计常广泛用于筛选实验. 8 次试验的 2^{4-1} 设计, 以及有 6 个因子、7 个因子或 8 个因子的 16 次试验的分式设计最为常用.

任一 2_{IV}^{k-p} 设计至少含有 $2k$ 次试验. 正好包含 $2k$ 次试验分辨度为Ⅳ的设计叫做**最小设计** (minimal design). 分辨度为Ⅳ的设计可以由分辨度为III的设计用**折叠**方法求得. 折叠一个 2_{III}^{k-p} 设计, 就是简单地把取反号的第二个分式设计和原来的分式设计相加. 于是, 在第一个分式设计中的单位列 I 的加号在第二个分式设计中就变号了, 而且第 $k+1$ 个因子与此列相联系. 这样得到一个 2_{IV}^{k+1-p} 分式析因设计. 对 2_{III}^{3-1} 设计, 可用表 8.27 说明这一方法. 容易证明, 得出的设计是定义关系为 $I = ABCD$ 的 2_{IV}^{4-1} 设计.

表 8.28 简要概括了 $N = 4$、8、16 和 32 个试验的 2^{k-p} 分式析因设计. 可以看到尽管 16 个试验分辨度为Ⅳ的设计可用于 $6 \leqslant k \leqslant 8$ 个因子, 但若有 9 个或更多的因子, 在 2^{9-p} 类型的设计中分辨度为Ⅳ的最小设计为 2^{9-4}, 也需要 32 个试验. 因为这试验次数较多, 所以许多实验者想要更少的设计. 考虑到分辨度为Ⅳ的设计至少要有 $2k$ 个试验, 所以一个九因子分辨度为Ⅳ的设计至少有 18 个试验. 恰有 $N = 18$ 个试验的设计可以用构造最优设计的算法来建立. 表 8.29 列出的设计是用这种算法与 D 最优设计准则构造的. 此准则选择设计点使得相应的回归

模型系数的方差最小. (最优设计与最优设计标准将在第 11 章更详细地进行讨论.) 表 8.29 中设计的别名关系 (只针对主效应与二因子交互作用) 是:

表 8.27 一个由折叠得到的 2_{IV}^{4-1} 设计

D I	A	B	C
	初始的 $2_{\text{III}}^{3-1}I = ABC$		
+	−	−	+
+	+	−	−
+	−	+	−
+	+	+	+
	取反号的第二个 2_{III}^{3-1} 设计		
−	+	+	−
−	−	+	+
−	+	−	+
−	−	−	−

$$
\begin{aligned}
[A] &= A, [B] = B, [C] = C, [D] = D, [E] = E, [F] = F, [G] = G, [H] = H, [J] = J \\
[AB] &= AB - 0.429BC - 0.429BD - 0.429BE + 0.429BF - 0.143BG + 0.429BH \\
&\quad - 0.429BJ + 0.571CG - 0.571CH + 0.571DG + 0.571DJ - 0.571EF \\
&\quad + 0.571EG - 0.571FG - 0.571GH + 0.571GJ \\
[AC] &= AC - 0.143BC - 0.143BD - 0.143BE + 0.143BF + 0.286BG - 0.857BH \\
&\quad - 0.143BJ - 0.143CG + 0.143CH + DF - 0.143DG - 0.143DJ + 0.143EF \\
&\quad - 0.143EG - EJ + 0.143FG - 0.857GH - 0.143GJ \\
[AD] &= AD - 0.143BC - 0.143BD - 0.143BE + 0.143BF + 0.286BG + 0.143BH \\
&\quad + 0.857BJ + CF - 0.143CG + 0.143CH - 0.143DG - 0.143DJ + 0.143EF \\
&\quad - 0.143EG + EH + 0.143FG + 0.143GH + 0.857GJ \\
[AE] &= AE - 0.143BC - 0.143BD - 0.143BE - 0.857BF + 0.286BG + 0.143BH \\
&\quad - 0.143BJ - 0.143CG + 0.143CH - CJ - 0.143DG + DH - 0.143DJ \\
&\quad + 0.143EF - 0.143EG - 0.857FG + 0.143GH - 0.143GJ \\
[AF] &= AF + 0.143BC + 0.143BD - 0.857BE - 0.143BF - 0.286BG - 0.143BH \\
&\quad + 0.143BJ + CD + 0.143CG - 0.143CH + 0.143DG + 0.143DJ - 0.143EF \\
&\quad - 0.857EG - 0.143FG - 0.143GH + 0.143GJ - HJ \\
[AG] &= AG + 0.571BC + 0.571BD + 0.571BE - 0.571BF - 0.143BG - 0.571BH \\
&\quad + 0.571BJ - 0.429CG - 0.571CH - 0.429DG + 0.571DJ - 0.571EF \\
&\quad - 0.429EG + 0.429FG + 0.429GH - 0.429GJ \\
[AH] &= AH - 0.857BC + 0.143BD + 0.143BE - 0.143BF - 0.286BG - 0.143BH
\end{aligned}
$$

$$+ 0.143BJ - 0.857CG - 0.143CH + DE + 0.143DG + 0.143DJ - 0.143EF$$
$$+ 0.143EG - 0.143FG - FJ - 0.143GH + 0.143GJ$$
$$[AJ] = AJ - 0.143BC + 0.857BD - 0.143BE + 0.143BF + 0.286BG + 0.143BH$$
$$- 0.143BJ - CE - 0.143CG + 0.143CH + 0.857DG - 0.143DJ + 0.143EF$$
$$- 0.143EG + 0.143FG - FH + 0.143GH - 0.143GJ$$

表 8.28　2^{k-p} 系统中的实用析因与分式析因设计，表中的数是实验中的因子个数

设计类型	实验个数			
	4	8	16	32
完全析因	2	3	4	5
$\frac{1}{2}$ 分式	3	4	5	6
分辨度为IV的分式	—	4	$6 \sim 8$	$7 \sim 16$
分辨度为III的分式	3	$5 \sim 7$	$9 \sim 15$	$17 \sim 31$

表 8.29　一个有 $k = 9$ 个因子 18 个试验分辨度为IV的最小设计

A	B	C	D	E	F	G	H	J
−	−	−	+	−	+	−	+	+
+	+	+	+	−	+	+	−	+
−	+	+	+	+	−	−	−	+
+	−	+	−	−	−	−	+	+
+	−	+	+	+	+	−	+	−
−	−	+	−	−	+	−	−	−
+	−	−	−	+	+	−	−	+
+	+	+	−	+	−	+	−	−
+	+	−	−	−	+	−	+	−
−	+	+	−	+	+	+	+	+
−	−	−	−	+	−	−	+	−
−	+	−	+	+	+	+	−	−
+	−	−	−	−	+	+	+	−
−	+	−	−	−	−	+	−	+
+	+	−	+	+	−	+	+	+
−	+	+	+	−	−	+	+	−
−	−	+	+	+	−	+	−	+
+	−	−	+	−	−	−	−	−

可以看出，在任一个分辨度为IV的设计中，可以估计出不受制于任意二因子交互作用的主效应，且二因子交互作用互为别名. 不过，二因子交互作用效应有部分混杂 (例如，BC 在不止一个别名链中出现). 18 次试验的设计中，二因子交互作用间的别名关系远比标准 32 次试验的 2^{9-4} 设计中的别名关系复杂 [根据附录表 X (p)]. 而且，主效应与交互作用的回归模型系数的标准误是 0.24σ，而在标准 32 次试验的 2^{9-4} 设计中是 0.18σ，所以在参数估计方面，18 次设计给出的精度不如标准 32 次设计. 最后，标准 2^{9-4} 设计是正交设计，而 18 次设计不是正交设计. 这使得 18 次设计的模型系数间有相关性，从而增大模型系数的标准误.

当因子有 $k = 6$ 或 7 个时，也想要构造分辨度为IV的最小设计来替代标准分辨度为IV的设计. 表 8.30 中列出 6 因子 12 个试验的分辨度为IV的最小设计. 该设计的别名关系为 (忽略三因子及更高阶交互作用):

$$[A] = A,\quad [B] = B,\quad [C] = C,\quad [D] = D,\quad [E] = E,\quad [F] = F$$

$$\begin{aligned}[AB] =& AB - 0.2BC + 0.6BD - 0.2BE - 0.6BF + 0.4CD - 0.8CE - 0.4CF \\ & + 0.4DE - 0.4DF - 0.4EF \\ [AC] =& AC + 0.2BC + 0.4BD - 0.8BE - 0.4BF + 0.6CD - 0.2CE - 0.6CF \\ & - 0.4DE + 0.4DF + 0.4EF \\ [AD] =& AD + 0.4BC - 0.2BD + 0.4BE - 0.8BF + 0.2CD - 0.4CE + 0.8CF \\ & + 0.2DE - 0.2DF + 0.8EF \\ [AE] =& AE - 0.8BC + 0.4BD + 0.2BE - 0.4BF - 0.4CD - 0.2CE + 0.4CF \\ & + 0.6DE + 0.4DF - 0.6EF \\ [AF] =& AF - 0.4BC - 0.8BD - 0.4BE - 0.2BF + 0.8CD + 0.4CE + 0.2CF \\ & + 0.8DE + 0.2DF + 0.2EF\end{aligned}$$

表 8.30 一个有 $k = 6$ 个因子 12 个试验分辨度为IV的最小设计

A	B	C	D	E	F
−	+	−	−	−	−
−	−	+	−	−	+
+	+	−	+	+	−
+	+	−	−	−	+
−	−	−	−	+	+
+	−	−	+	−	−
−	+	+	−	+	+
−	−	+	+	+	−
+	−	+	−	+	−
+	−	+	+	+	+
+	+	+	+	−	−
−	+	−	+	−	+

仍要注意, 为了将试验次数从 16 次降为 12 次, 实验者要将二因子交互作用的别名关系更加复杂化. 相对于标准设计而言, 在模型系数估计的精度方面也有损失.

这些分辨度IV的最小设计是非正规分式析因设计的一些例子. Design-Expert 包含了因子个数 $5 \leqslant k \leqslant 30$ 的一系列这类设计. 在主效应最受关注而二因子交互作用不能完全忽略的筛选问题中, 这些设计是标准 2_{IV}^{k-p} 析因设计的非常有用的替代. 若二因子交互作用是重要的, 为了确定哪些交互作用是重要的, 则必须追加实验.

8.6.2 分辨度为IV的设计的序贯实验

因为分辨度为IV的设计常用于筛选实验, 在实施与分析了最初的实验后, 经常需要用追加实验来完全分辨所有效应. 我们在 8.5.2 节中讨论了分辨度为III的设计的情况, 并引入了**折叠**作为序贯实验的方案. 在分辨度为III的场合, 主效应与二因子交互作用混杂, 所以折叠的目的是将主效应与二因子交互作用分离. 也有可能通过折叠分辨度为IV的设计来分离相互混杂的二因子交互作用.

Montgomery and Runger(1996) 注意到实验者折叠分辨度为Ⅳ的设计时的几个目标, 就是

(1) 将尽可能多的二因子交互作用别名链打破;

(2) 打破某个特殊的二因子交互作用别名链, 或

(3) 打破与某个特殊因子有关的二因子交互作用别名链.

然而, 折叠分辨度为Ⅳ的设计时必须要小心. 我们用于分辨度III的设计时的完全折叠规则, 即简单地实施全体符号都取相反号的另一个分式设计, 对于分辨度为Ⅳ的设计就不奏效了. 若该规则用于分辨度为Ⅳ的设计, 结果会得到与原先设计完全一样只是试验排序不同的设计. 试试看! 用表 8.9 中的 2_{IV}^{6-2} 设计, 你把试验矩阵中的符号取相反号后, 看看会发生什么?

折叠分辨度为Ⅳ的设计的最简单的方法是转换原先设计中某个变量的符号. 这种单因子折叠使得所有包含该因子的二因子交互作用的符号也发生转换, 从而被分离, 这样完成了上述的第 3 个目标.

为了说明如何完成分辨度为Ⅳ的设计的单因子折叠, 考虑表 8.31 中的 2_{IV}^{6-2} 设计 (试验按标准顺序排列, 并非按实施时的顺序). 此实验用于研究关于硅晶片光阻镀层厚度的 6 个因子的效应. 设计因子是 $A=$ 旋转速度, $B=$ 加速度, $C=$ 所用光阻剂的体积, $D=$ 旋转时间, $E=$ 光阻剂黏度, $F=$ 排气率. 此设计的别名关系见表 8.8. 效应的半正态概率图见图 8.25. 可以看到最大的效应是 A, B, C, E, 因为这些效应与三因子交互作用或更高阶的交互作用混杂, 所以假定这些效应是实际效应是合乎逻辑的. 然而, 关于 $AB+CE$ 别名链的效应估计也偏大. 除非还有其他工序知识或工程信息可提供, 要不然我们不知道这究竟是交互作用效应 AB, CE 还是两个都是.

表 8.31 硅晶片光阻镀层实验原先的 2_{IV}^{6-2} 设计

A	B	C	D	E	F	
速度 (RPM)	加速度	体积 (cc)	时间 (Sec)	光阻剂黏度	排气率	厚度 (Mil)
−	−	−	−	−	−	4 524
+	−	−	−	+	−	4 657
−	+	−	−	+	+	4 293
+	+	−	−	−	+	4 516
−	−	+	−	+	+	4 508
+	−	+	−	−	+	4 432
−	+	+	−	−	−	4 197
+	+	+	−	+	−	4 515
−	−	−	+	−	+	4 521
+	−	−	+	+	+	4 610
−	+	−	+	+	−	4 295
+	+	−	+	−	−	4 560
−	−	+	+	+	−	4 487
+	−	+	+	−	−	4 485
−	+	+	+	−	+	4 195
+	+	+	+	+	+	4 510

通过建立一个新的 2_{IV}^{6-2} 分式析因设计, 并改变因子 A 的符号, 可构造出一个折叠设计. 加上了折叠试验后的完全设计被列在 (按标准顺序) 表 8.32 中. 可以看到试验分成两个区组, 来自原先表 8.31 2_{IV}^{6-2} 设计的试验在区组 1 中, 折叠试验在区组 2 中. 从组合试验得到的效应估计 (忽略三因子交互作用或更多因子的交互作用):

$$
\begin{aligned}
&[A]=A && [AE]=AE\\
&[B]=B && [AF]=AF\\
&[C]=C && [BC]=BC+DF\\
&[D]=D && [BD]=BD+CF\\
&[E]=E && [BE]=BE\\
&[F]=F && [BF]=BF+CD\\
&[AB]=AB && [CE]=CE\\
&[AC]=AC && [DE]=DE\\
&[AD]=AD && [EF]=EF
\end{aligned}
$$

可以看到, 现在所有包含因子 A 的二因子交互作用与其他二因子交互作用分离了. 而且, AB 也不再与 CE 混杂了. 从组合的设计得到的效应的半正态概率图画在图 8.26 中. 显然交互作用 CE 是显著的.

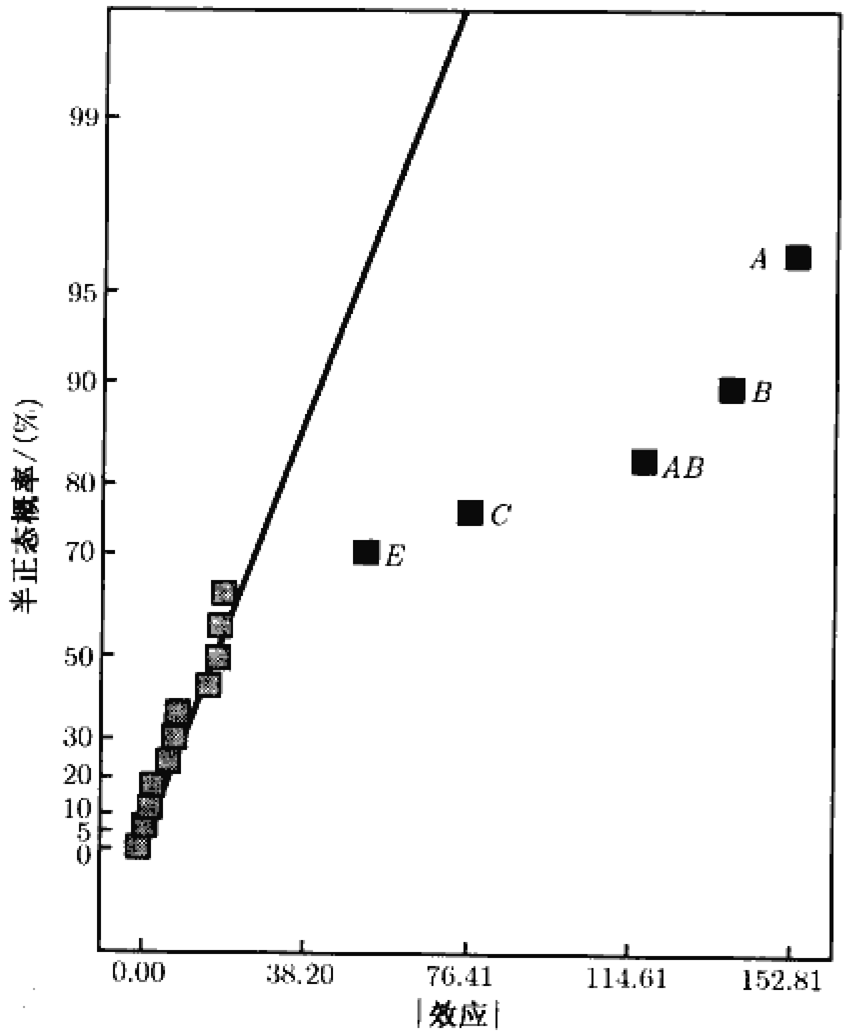


图 8.25 表 8.31 中硅晶片光阻镀层实验原先的效应的半正态概率图

容易证明, 表 8.32 中的完全折叠设计可以估计前面列出的 6 个主效应和 12 个二因子交互作用别名链, 以及其他 12 个包括高阶交互作用与区组效应的别名链. 原先分式设计的生成元是 $E=ABC$ 和 $F=BCD$, 因为我们改变了 A 列的符号来建立折叠设计, 所以关于第二组 16 个试验生成元是 $E=-ABC$ 和 $F=BCD$. 因为只有一个单词是一样的 $(L=1, U=1)$, 所以组合的设计只有一个生成元 (它是一个 1/2 分式设计), 组合的设计的生成元是 $F=BCD$. 而且, 因为 $ABCE$ 在区组 1 中为正, 在区组 2 中为负, 所以 $ABCE$ 加上它的别名 $ADEF$ 都与区组混杂.

表 8.32　硅晶片光阻镀层实验的整个折叠设计

		A	B	C	D	E	F	
标准顺序	区组	速度 (转/分)	加速度	体积 (cc)	时间 (sec)	光阻剂黏度	排气率	厚度 (mil)
1	1	−	−	−	−	−	−	4 524
2	1	+	−	−	−	+	−	4 657
3	1	−	+	−	−	+	+	4 293
4	1	+	+	−	−	−	+	4 516
5	1	−	−	+	−	+	+	4 508
6	1	+	−	+	−	−	+	4 432
7	1	−	+	+	−	−	−	4 197
8	1	+	+	+	−	+	−	4 515
9	1	−	−	−	+	−	+	4 521
10	1	+	−	−	+	+	+	4 610
11	1	−	+	−	+	+	−	4 295
12	1	+	+	−	+	−	−	4 560
13	1	−	−	+	+	+	−	4 487
14	1	+	−	+	+	−	−	4 485
15	1	−	+	+	+	−	+	4 195
16	1	+	+	+	+	+	+	4 510
17	2	+	−	−	−	−	−	4 615
18	2	−	−	−	−	+	−	4 445
19	2	+	+	−	−	+	+	4 475
20	2	−	+	−	−	−	+	4 285
21	2	+	−	+	−	+	+	4 610
22	2	−	−	+	−	−	+	4 325
23	2	+	+	+	−	−	−	4 330
24	2	−	+	+	−	+	−	4 425
25	2	+	−	−	+	−	+	4 655
26	2	−	−	−	+	+	+	4 525
27	2	+	+	−	+	+	−	4 485
28	2	−	+	−	+	−	−	4 310
29	2	+	−	+	+	+	−	4 620
30	2	−	−	+	+	−	−	4 335
31	2	+	+	+	+	−	+	4 345
32	2	−	+	+	+	+	+	4 305

检查包含原先 16 个试验的设计中二因子交互作用的别名链, 以及整个折叠设计的二因子交互作用的别名链, 可以发现一些麻烦的信息. 在原先分辨度为Ⅳ的分式设计中, 6 个别名链中每个二因子交互作用都与其他二因子交互作用混杂, 且其中的一个别名链中有 3 个二因子交互作用 (根据表 8.8). 因此, 7 个自由度可用于估计二因子交互作用. 在完全折叠设计中, 有 9 个二因子交互作用估计时与其他二因子交互作用独立, 而且有三个包含两个二因子交互作用的别名链, 这使得 12 个自由度可用于估计二因子交互作用. 换种说法, 我们增加了 16 次试验, 却只增加了 5 个自由度用于估计二因子交互作用. 这样就没有充分利用实验资源.

幸运的是, 还有不同于完全折叠的其他方案. 采用**部分折叠**(或**半折叠**), 我们只采用完全折叠的一半试验次数, 对于旋转镀层实验来讲只需要 8 个试验. 进行部分折叠设计的步骤如下.

(1) 用通常的办法根据原设计构造一个单因子折叠设计, 即改变感兴趣的二因子交互作用的任一因子的符号.

(2) 选取被选因子取高水平或低水平的那些试验, 选出折叠设计的一半试验. 选择你认为能得到最理想响应值的那个水平, 通常是个好主意.

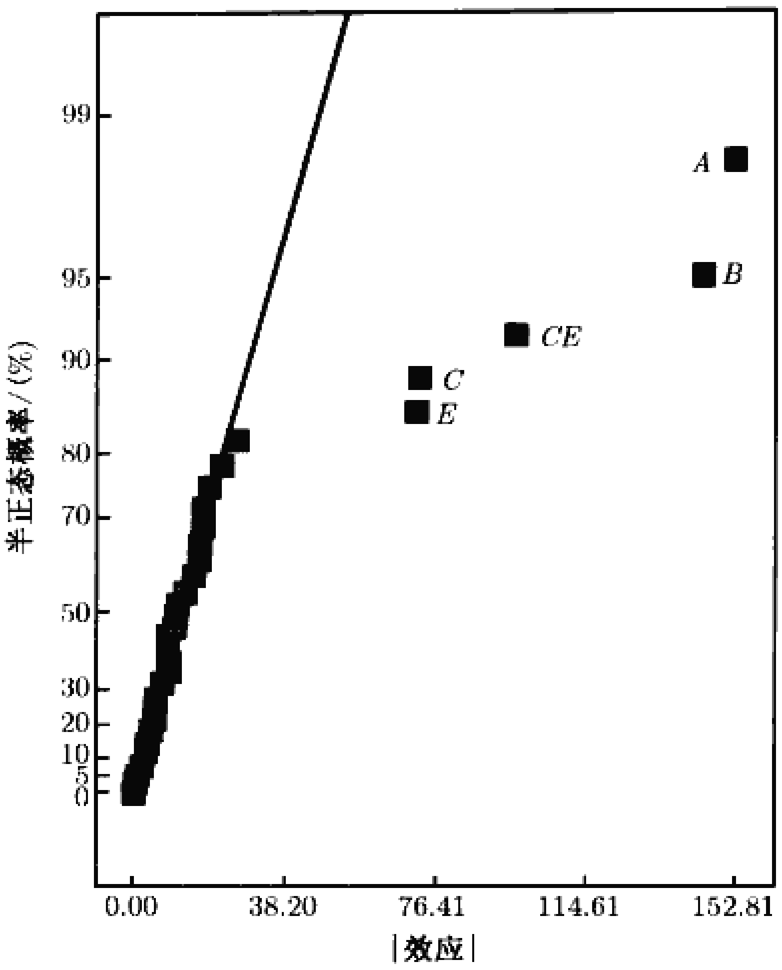


图 8.26 表 8.32 中硅晶片光阻镀层实验的部分折叠设计效应的半正态概率图

表 8.33 是关于旋转镀层实验的部分折叠设计. 注意我们选择 A 是低水平的那些试验, 这是

表 8.33 硅晶片光阻镀层实验的部分折叠设计

		A	B	C	D	E	F	
标准顺序	区组	速度 (转/分)	加速度	体积 (cc)	时间 (sec)	光阻剂黏度	排气率	厚度 (mil)
1	1	−	−	−	−	−	−	4 524
2	1	+	−	−	−	+	−	4 657
3	1	−	+	−	−	+	+	4 293
4	1	+	+	−	−	−	+	4 516
5	1	−	−	+	−	+	+	4 508
6	1	+	−	+	−	−	+	4 432
7	1	−	+	+	−	−	−	4 197
8	1	+	+	+	−	+	−	4 515
9	1	−	−	−	+	−	+	4 521
10	1	+	−	−	+	+	+	4 610
11	1	−	+	−	+	+	−	4 295
12	1	+	+	−	+	−	−	4 560
13	1	−	−	+	+	+	−	4 487
14	1	+	−	+	+	−	−	4 485
15	1	−	+	+	+	−	+	4 195
16	1	+	+	+	+	+	+	4 510
17	2	−	−	−	−	+	−	4 445
18	2	−	+	−	−	−	+	4 285
19	2	−	−	+	−	−	+	4 325
20	2	−	+	+	−	+	−	4 425
21	2	−	−	−	+	+	+	4 525
22	2	−	+	−	+	−	−	4 310
23	2	−	−	+	+	−	−	4 335
24	2	−	+	+	+	+	+	4 305

因为在原先的 16 次试验中 (表 8.31), A 取低水平时得到了较薄的光阻镀层 (这是该实验需要的). (在对原先 16 次试验的分析中, A 的效应估计是正的, 这也表明 A 取低水平时产生所需要的结果.)

部分折叠产生的别名关系 (忽略三因子交互作用或更多因子的交互作用) 是:

$$
\begin{array}{ll}
[A]=A & [AE]=AE \\
[B]=B & [AF]=AF \\
[C]=C & [BC]=BC+DF \\
[D]=D & [BD]=BD+CF \\
[E]=E & [BE]=BE \\
[F]=F & [BF]=BF+CD \\
[AB]=AB & [CE]=CE \\
[AC]=AC & [DE]=DE \\
[AD]=AD & [EF]=EF
\end{array}
$$

可以看到, 有 12 个自由度能用于估计二因子交互作用, 与完全折叠一样多. 而且, AB 不再与 CE 混杂. 通过部分折叠得到的效应的半正态概率图见图 8.27. 和完全折叠设计一样, 认为 CE 是显著的二因子交互作用.

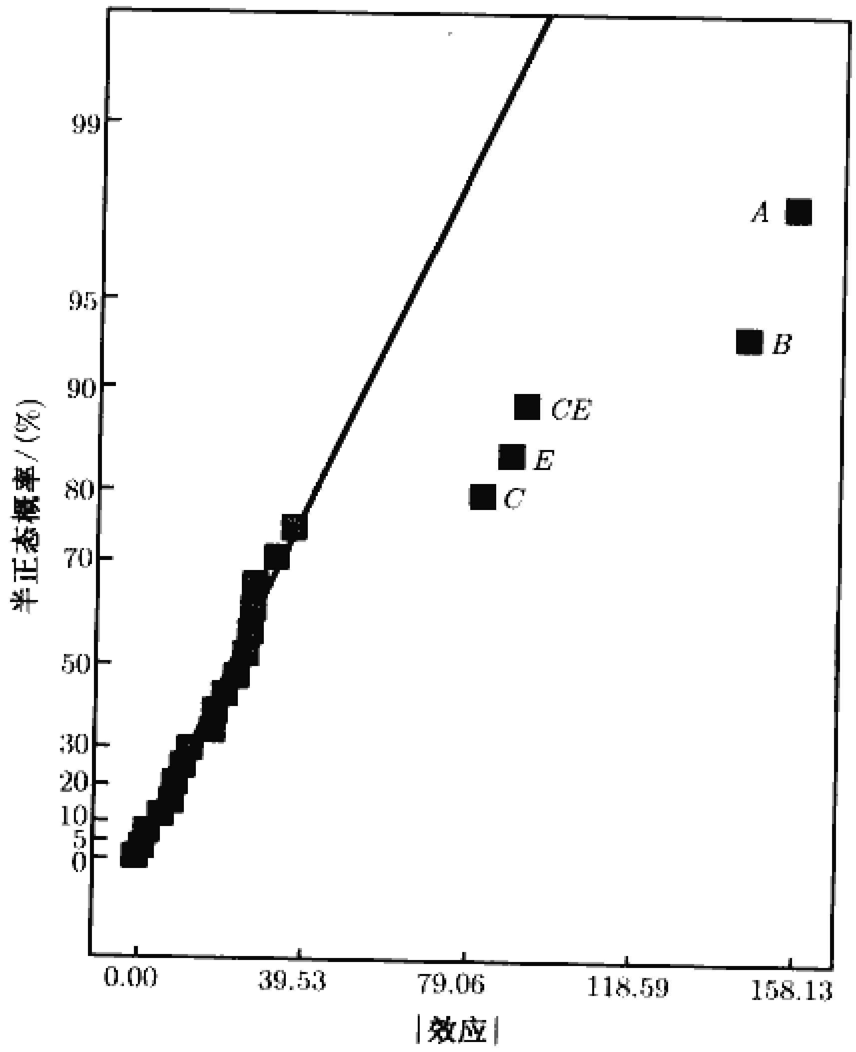


图 8.27　表 8.32 中硅晶片光阻镀层实验的效应的半正态概率图

部分折叠技术对分辨度为Ⅳ的设计非常有用, 通常能够有效使用实验资源. 分辨度为Ⅳ的设计总是可以给出主效应好的估计 (假定三因子交互作用被忽略), 而且需要分离的二因子交互作

用的个数并不很大. 对分辨度为Ⅳ的设计的部分折叠能够估计二因子交互作用的个数与完全折叠时相等. 部分折叠的一个缺点在于它是非正规的分式设计, 因此不是正交设计. 这导致参数估计是相关的, 并使得效应或回归模型系数的标准误增大. 例如, 在旋转镀层实验的部分折叠设计中, 回归模型系数的标准误从 0.20σ 增加到 0.25σ, 而采用完全折叠设计, 它是正交的, 模型系数的标准误为 0.18σ. 关于部分折叠的更多内容, 参见 Mee and Peralta(2000) 以及本章的补充材料.

8.6.3 分辨度为 V 的设计

分辨度为 V 的设计是这样的分式析因设计, 它的主效应与二因子交互作用不能以另外的主效应和二因子交互作用作为它的别名. 因此, 这些设计是非常强的设计, 当三因子交互作用和更高阶的交互作用被忽略时, 这些设计可以单独地估计出所有的主效应与二因子交互作用. 一个分辨度为 V 的设计的定义关系中最短的词必须有 5 个字母. 生成关系为 $I = ABCDE$ 的 2^{5-1} 设计可能是应用最广的分辨度为 V 的设计. 它只用 16 次试验, 就能研究 5 个因子, 并估计所有 5 个主效应和 10 个二因子交互作用. 我们在例 8.2 中曾说明过这种设计的使用.

当因子个数 $k = 6$ 时, 分辨度至少为 V 的最小设计是有 32 次试验的 2_{VI}^{6-1} 设计, 它的分辨度为Ⅵ. 当因子个数 $k = 7$ 时, 最小设计则是 64 次试验的 2_{V}^{7-1} 设计, 它的分辨度为Ⅶ. 因子个数 $k = 8$ 时, 最小设计则是 64 次试验的 2_{V}^{8-2} 设计. 对于因子个数 $k \geqslant 9$ 时, 所有这类设计都至少需要 128 次试验. 这些都是非常大的设计, 所以统计学家一直想要找出能达到所需分辨度较小的替代设计. Mee(2004) 给出这个课题的综述. 非正规分式设计非常有用. 表 8.34 是因子个数 $k = 6$ 且试验次数 $N = 22$ 的非正规二水平分式设计. 此设计与 2_{VI}^{6-1} 设计一样能估

表 8.34 因子个数 $k = 6$ 分辨度为V的二水平分式设计

A	B	C	D	E	F
+	−	−	−	+	−
+	−	+	−	+	+
−	+	+	−	−	−
−	−	−	−	+	+
+	+	−	+	+	+
+	+	−	+	−	+
+	−	−	+	−	+
+	+	−	−	−	+
−	−	−	+	+	−
−	−	−	−	−	−
+	−	+	+	+	−
−	+	−	−	+	−
+	+	+	+	−	+
+	−	+	−	−	−
−	−	+	−	−	+
+	+	−	+	−	−
−	−	+	+	+	+
−	−	+	+	−	−
−	+	−	+	−	+
+	+	+	−	+	−
−	+	+	+	+	−
−	+	+	−	+	+

计所有主效应与二因子交互作用, 但减少了 10 次试验. 然而表 8.34 中的设计不是正交的, 而且这也影响了对效应与回归系数估计的精度. 对于表 8.34 中的设计, 回归模型系数的标准误从 0.26σ 增加到 0.29σ, 而在 2_{VI}^{6-1} 设计中, 相应的标准误为 0.18σ.

作为最后的例子, 表 8.35 给出了因子个数 $k=8$ 且试验次数 $N=38$ 的非正规二水平分式设计. 此设计与 2_{V}^{8-2} 设计一样能估计所有主效应与二因子交互作用, 但少了 26 次试验. 然而此设计的非正交性影响了对效应与回归系数估计的精度. 对于表 8.35 中的设计, 回归模型系数的标准误从 0.18σ 增加到 0.26σ, 而在 2_{V}^{8-2} 设计中, 相应的标准误为 0.13σ.

表 8.35 因子个数 $k=8$ 分辨度为 V 的二水平分式设计

A	B	C	D	E	F	G	H
−	−	+	−	+	−	+	−
−	+	+	−	+	+	−	+
−	+	+	+	−	−	−	−
−	−	−	−	−	−	+	−
+	+	−	−	+	−	+	−
+	−	+	+	+	+	+	−
−	−	−	+	+	+	−	−
+	−	−	+	+	−	−	−
−	−	+	−	−	+	+	+
+	+	+	−	+	−	−	+
+	+	+	−	−	+	−	+
+	+	−	+	−	−	+	+
−	+	−	+	+	−	+	−
+	−	−	−	−	−	+	+
+	+	−	+	+	+	+	−
−	+	−	−	−	+	+	+
+	−	+	+	−	+	−	−
+	−	+	−	+	+	−	−
−	−	−	−	−	+	−	−
−	+	+	+	−	+	+	+
−	+	+	−	−	−	+	+
−	+	−	+	+	−	−	+
+	+	+	+	−	−	+	−
−	−	−	+	+	−	+	+
+	−	+	+	−	−	+	+
+	−	+	−	−	−	−	−
+	−	−	+	−	+	+	+
+	+	−	−	+	+	−	+
+	+	+	−	+	+	+	+
+	+	−	−	−	−	−	−
−	+	+	−	−	+	+	−
−	−	−	−	+	+	+	−
−	−	−	+	−	−	−	+
+	−	+	+	+	+	−	+
+	+	+	+	+	+	−	−
−	−	+	+	+	−	−	+
−	−	−	−	+	−	−	+
−	+	−	+	−	+	−	−

尽管损失了估计的精度, 当实验资源非常缺乏的时候, 这些非正规设计仍然非常有用. Design-Expert 包含了因子个数 $6 \leqslant k \leqslant 30$ 的一系列这类设计. 这些设计是采用前面讨论过的关于最小分辨度为Ⅳ的最小设计的相同方法构造的.

8.7 超饱和设计

饱和设计的定义是：试验次数为 N, 且因子或实验变量的个数为 $k = N - 1$ 的分式析因设计. 最近几年, 有相当多的兴趣是关注于**超饱和设计**的构造及其在筛选实验中的应用. 在超饱和设计中, 变量个数 $k > N - 1$, 而且这些设计一般包含了比试验次数多得多的变量. 应用超饱和设计的想法最初是由 Satterthwaite(1959) 提出的. 他建议随机生成这些设计. 在对此文的广泛讨论中, 当时实验设计方面的一些权威, 包括 Jack Youden, George Box, J. Stuart Hunter, William Cochran, John Tukey, Oscar Kempthorne 和 Frank Anscombe, 都批评了随机平衡设计. 结果, 超饱和设计在随后 30 年里很少受到注意. 有一个例外是 Booth and Cox(1962) 提出的系统超饱和设计. 他们的设计不是随机生成的, 这一点明显不同于 Satterthwaite(1959) 所提出的. Booth 和 Cox 用基本的计算机搜索方法产生他们的设计. 他们还建立了评价超饱和设计的基本原则.

Lin(1993) 再次研究了超饱和设计概念, 引发了关于此课题的更多研究. 许多作者也提出了构造超饱和设计的方法. Lin(2000) 中有很好的综述. 大多数构造设计的方法局限于简单的计算机搜索程序 [见 Lin(1995), Li and Wu(1995), Holcomb, and Carlyle(2002)]. 其他人还提出了基于最优设计构造技术的方法 (在第 11 章我们将讨论最优设计).

另一种构造超饱和设计的方法是基于现有的正交设计的框架. 包括采用 1/2 的 Hadamard 矩阵 [Lin(1993)], 以及计算某些 Hadamard 矩阵的二因子交互作用 [Wu(1995)]. Hadamard 矩阵是一种元素为 -1 或 1 的正交方阵. 当实验中因子的个数超过了试验次数, 设计矩阵就不可能是正交的. 因此, 因子效应的估计就不独立. 具有支配性因子的实验可能会影响或歪曲另一个因子的作用. 构造超饱和设计是为了最小化因子间的非正交性.

基于 Hadamard 矩阵 1/2 分式的超饱和设计很容易构造. 表 8.36 是有 11 个因子和 12 次试验的 Plackett-Burman 设计. 它也是一个 Hadamard 矩阵设计. 该表中, 设计已按最后一列 (因子 11 或因子 L) 排序. 有时称此列为**分支列**. 现只保留该设计中 L 列取正号的那些试验, 并在这些试验中删去 L 列. 得到的设计是有 $k = 10$ 个因子和 6 次试验的超饱和设计. 我们也可以用 L 列取负号的那些试验. 这种方法总可以得到有 $k = N - 2$ 个因子和 $N/2$ 次试验的超饱

表 8.36 由 12 次试验的 Hadamard 矩阵 (Plackett-Burman) 设计得到的超饱和设计

试验	I	因子 1(A)	因子 2(B)	因子 3(C)	因子 4(D)	因子 5(E)	因子 6(F)	因子 7(G)	因子 8(H)	因子 9(J)	因子 10(K)	因子 11(L)
1	+	−	+	+	+	−	+	+	−	+	−	−
2	+	−	−	−	−	−	−	−	−	−	−	−
3	+	+	−	−	−	+	+	+	−	+	+	−
4	+	−	−	+	+	+	−	+	+	−	+	−
5	+	+	+	+	−	+	+	−	+	−	−	−
6	+	+	+	−	+	−	−	−	+	+	+	−
7	+	−	+	+	−	+	−	−	−	+	+	+
8	+	+	+	−	+	+	−	+	−	−	−	+
9	+	+	−	+	−	−	−	+	+	+	−	+
10	+	+	−	+	+	−	+	−	−	−	+	+
11	+	−	−	−	+	+	+	−	+	+	−	+
12	+	−	+	−	−	−	+	+	+	−	+	+

和设计. 若感兴趣的因子不足 $N-2$ 个, 则可从完全设计中去掉一些列.

超饱和设计通常用回归模型拟合的方法进行分析, 例如向前选择 (见第 10 章). 这种方法中, 变量一次一个地被选入模型直到没有其他变量对解释响应有明显作用为止. Abraham, Chipman and Vijayan(1999) 以及 Holcomb, Montgomery, and Carlyle(2003) 已研究了超饱和设计的分析方法. 一般来讲, 这些设计可以经受大的第 I 类和第 II 类错误, 但有些分析方法主要强调第 I 类错误, 从而第 II 类错误率会是中等的. 在因子筛选时, 比起把不积极的因子误认为是积极因子来讲, 不漏掉积极的因子通常更为重要. 所以第 I 类错误不如第 II 类错误更关键. 然而, 因为两种错误率可能都比较大, 应用超饱和设计的原则应该是剔除大部分的不积极的因子, 而并不需要清楚识别出那些少数重要的或积极的因子. Holcomb, Montgomery, and Carlyle(2003) 发现某些超饱和设计在第 I 类和第 II 类错误方面比其他超饱和设计有更好的表现. 一般地, 由搜索算法得到的超饱和设计要优于那些由标准正交设计构造的超饱和设计.

超饱和设计还没有被广泛使用. 不过, 当实验所研究的系统有很多变量但其中只有极少数会有大的效应时, 超饱和设计是有趣而且有用的方法.

8.8 小　结

本章引进了 2^{k-p} 分式析因设计. 我们已经强调了在筛选实验中使用这些设计, 以便快速而有效地识别出积极的因子子集, 并且提供关于交互作用的一些信息. 这些设计的投影性质, 在大多数情况下使得它能够更为详尽地检测那些积极因子. 对在实验中识别为可能重要的交互作用, 可通过非常有效的折叠法把设计序贯地组装起来, 以求得关于交互作用的附加信息.

在实践中, $N=4$、8、16 和 32 个试验的 2^{k-p} 分式析因设计有很高的使用价值. 表 8.28 列出了这些设计, 指出了: 做各类筛选实验时, 按照有多少个因子应使用哪一个设计. 例如, 做 16 个试验的设计, 四因子的应是完全析因设计, 五因子的是 1/2 分式设计, 6~8 个因子的是分辨度为IV的分式设计, $9\sim15$ 因子的是分辨度为III的设计. 所有这些设计都可以用本章讨论的方法来构造, 而且其中许多设计的别名结构已列在附录的表 X 中.

8.9 思　考　题

8.1 在思考题 6.7 所述的化学过程生产改进实验中, 假设只能做 2^4 设计的 1/2 分式设计. 构造这一设计, 并用重复 I 的数据进行统计分析.

8.2 假设在思考题 6.15 中, 只能做 2^4 设计的 1/2 分式设计. 构造这一设计, 并用重复 I 的数据进行统计分析.

8.3 考虑例 6.1 所述的等离子蚀刻实验. 假定只可以做 1/2 分式设计. 建立此设计并分析数据.

8.4 思考题 6.24 提到一个关于集成电路生产过程的过程改进研究. 假定只可以做 8 次试验. 建立一恰当的 2^{5-2} 设计, 并求出别名结构. 用思考题 6.21 的数据作为此设计的观测值, 估计因子效应. 你能得出什么结论?

8.5 (续思考题 8.4)假设你已做完了思考题 8.4 中的 2^{5-2} 设计的 8 个试验. 为了识别出感兴趣的因子效应, 需附加什么试验? 在组合的设计中, 别名关系是什么?

8.6 R. D. Snee ("Experimenting with a Large Number of Variables", *Experiment in Industry: Design, Analysis and Interpretation of Results*, R. D. Snee, L. B. Hare and J. B. Trout, Editors, ASQC,

1985) 描述了一个实验, 是 $I = ABCDE$ 的 2^{5-1} 设计, 用来研究 5 个因子对一种化学产品的颜色的效应. 这 5 个因子是: $A=$ 溶剂/反应物、$B=$ 催化剂/反应物、$C=$ 温度、$D=$ 试剂纯度、$E=$ 试剂的 PH 值. 所得结果如下.

$e = -0.63$ $d = 6.79$ $a = 2.51$ $ade = 5.47$ $b = -2.68$ $bde = 3.45$ $abe = 1.66$

$abd = 5.68$ $c = 2.06$ $cde = 5.22$ $ace = 1.22$ $acd = 4.38$ $bce = -2.09$

$bcd = 4.30$ $abc = 1.93$ $abcde = 4.05$

(a) 画出效应的正态概率图, 哪些效应是积极的?

(b) 计算残差. 画出残差的正态概率图以及残差与拟合值的关系图. 评论这些图形.

(c) 如果有因子可忽略, 把 2^{5-1} 设计压缩成只是关于积极因子的完全析因设计. 评论所得的设计并解释结果.

8.7 在*Journal of Quality Technology* (Vol. 17, 1985, pp.198~206) 上, Pignatiello 和 Ramberg 的一篇论文描述了有重复的分式析因设计, 研究 5 个因子对汽车使用的簧片自由高度的效应. 这些因子为: $A=$ 炉温、$B=$ 加热时间、$C=$ 传递时间、$D=$ 保持时间、$E=$ 淬火油温. 得到的数据如下:

A	B	C	D	E	自由高度		
−	−	−	−	−	7.78	7.78	7.81
+	−	−	+	−	8.15	8.18	7.88
−	+	−	+	−	7.50	7.56	7.50
+	+	−	−	−	7.59	7.56	7.75
−	−	+	+	−	7.54	8.00	7.88
+	−	+	−	−	7.69	8.09	8.06
−	+	+	−	−	7.56	7.52	7.44
+	+	+	+	−	7.56	7.81	7.69
−	−	−	−	+	7.50	7.25	7.12
+	−	−	+	+	7.88	7.88	7.44
−	+	−	+	+	7.50	7.56	7.50
+	+	−	−	+	7.63	7.75	7.56
−	−	+	+	+	7.32	7.44	7.44
+	−	+	−	+	7.56	7.69	7.62
−	+	+	−	+	7.18	7.18	7.25
+	+	+	+	+	7.81	7.50	7.59

(a) 写出这个设计的别名结构, 这个设计的分辨度是几?

(b) 分析这些数据. 是什么因子影响了平均自由高度?

(c) 对每个试验计算自由高度的极差和标准差. 有迹象表明这些因子中的哪些因子影响自由高度的变异性吗?

(d) 分析此实验的残差, 并评论你的发现.

(e) 这是一个有 5 个因子 16 个试验的最好设计吗? 具体说就是, 你能否找出一个有 5 个因子 16 个试验的分式设计比这一设计具有更高的分辨度?

8.8 在*Industrial and Engineering Chemistry* 上的一篇论文 ("More on Planning Experiments to Increase Research Efficiency", 1970, pp. 60-65) 中, 用 2^{5-2} 设计研究 $A=$ 冷凝温度 $B=$ 材料 1 的量、$C=$ 溶剂量、$D=$ 凝结时间以及 $E=$ 材料 2 的量等因子对产量的影响. 结果如下

$e = 23.2$ $ad = 16.9$ $cd = 23.8$ $bde = 16.8$

$ab = 15.5$ $bc = 16.2$ $ace = 23.4$ $abcde = 18.1$

(a) 证实设计所用的生成元是 $I = ACE$ 和 $I = BDE$.

(b) 写出这个设计的完全定义关系和别名.

(c) 估计主效应.

(d) 写出方差分析表. 证明 AB 和 AD 交互作用可用作误差.

(e) 画出残差对拟合值的图形. 并画出残差的正态概率图. 评论这些结果.

8.9 考虑思考题 8.7 的簧片实验. 假设在制造簧片时因子 E (淬火油温) 非常难于控制. 为了尽量减少自由高度的变异性, 不管使用什么淬火油温, 你将怎样设置因子 A, B, C, D 的水平?

8.10 选取两个四因子交互作用作为独立生成元构造一个 2^{7-2} 设计. 写出此设计的完全别名结构. 列出方差分析表. 这个设计的分辨度是几?

8.11 考虑思考题 6.24 中的 2^5 设计. 假定只能做 1/2 分式设计实验. 而且, 取得 16 个观测值需要两天, 这就需用二区组的 2^{5-1} 混区设计. 构造此设计并分析这些数据.

8.12 分析思考题 6.26 中的数据, 把它作为来自一个 $I = ABCD$ 的 2_{IV}^{4-1} 设计. 把此设计投影到一个原先四因子的显著因子子集的完全析因设计中.

8.13 用 $I = -ABCD$ 重做思考题 8.12. 用备选分式设计是否改变了你对数据的解释?

8.14 将例 8.1 中的 2_{IV}^{4-1} 设计投影为因子 A 和 B 的一个有两次重复的 2^2 设计. 分析数据并写出结论.

8.15 构造一个 2_{III}^{5-2} 设计, 如果实施了此设计的完全折叠设计, 确定可被估计的效应.

8.16 构造一个 2_{III}^{6-3} 设计, 如果实施了此设计的完全折叠设计, 确定可被估计的效应.

8.17 考虑思考题 8.15 中的 2_{III}^{6-3} 设计. 若此设计的第二个分式设计是取因子 A 的反号试验, 确定可被估计的效应.

8.18 折叠表 8.19 中的 2_{III}^{7-4} 设计得到一个八因子设计. 证实所得设计是一个 2_{IV}^{8-4} 设计. 这是最小的设计吗?

8.19 折叠一个 2_{III}^{5-2} 设计得到一个六因子设计. 证实所得设计是一个 2_{IV}^{6-2} 设计. 将此设计和表 8.10 的 2_{IV}^{6-2} 设计进行比较.

8.20 一个工业工程师用库存系统的蒙特卡罗模拟模型进行了一次实验. 她的模型的独立变量是订货量 (A)、再订购点 (B)、装置成本 (C)、预订货价 (D)、搬运费率 (E). 为了节省计算机的时间, 她决定用 $I = ABD$ 和 $I = BCE$ 的 2_{III}^{5-2} 设计来研究这些因子. 所得结果为: $de = 95$, $ae = 134$, $b = 158$, $abd = 190$, $cd = 92$, $ac = 187$, $bce = 155$, $abcde = 185$.

(a) 证实给出的处理组合是正确的. 假定三因子交互作用和更高阶的交互作用可被忽略, 估计这些效应.

(b) 假定和第一个分式设计相加的第二个分式设计是, $ade = 136$, $e = 93$, $ab = 187$, $bd = 153$, $acd = 139$, $c = 99$, $abce = 191$, $bcde = 150$. 这第二个分式设计是怎样求得的? 将这些数据添加到原来的分式设计中, 估计效应.

(c) 假定进行了试验的分式设计是 $abc = 189$, $ce = 96$, $bcd = 154$, $acde = 135$, $abe = 193$, $bde = 152$, $ad = 137$, $(1) = 98$. 怎样求得这一分式设计? 将这些数据添加到原来的分式设计中, 估计效应.

8.21 构造一个 2^{5-1} 设计. 说明此设计是怎样实施在两个区组中的, 每区组有 8 个试验, 有主效应或二因子交互作用与区组相混杂吗?

8.22 构造一个 2^{7-2} 设计, 说明此设计是怎样实施在 4 个区组中的, 每区组有 8 个试验, 有主效应或二因子交互作用与区组相混杂吗?

8.23 2^k 的非正规分式设计 [John(1971)]. 考虑一个 2^4 设计. 必须估计 4 个主效应和 6 个二因子交互作用, 但不能进行完全 2^4 析因设计实验. 最大可能的区组大小含有 12 个试验. 这 12 个试验可以由 $I = \pm AB = \pm ACD = \pm BCD$ 定义的, 由 4 个 1/4 分式设计中的 3 个 (去掉主分式) 来完成. 假定可忽略三因子交互作用及更高阶的交互作用, 为了估计所需的效应, 应该怎样把除去主分式设计后

留下的 3 个 2^{4-2} 分式设计组合起来? 这设计可看作为一个 3/4 分式设计.

8.24 精炼过程使用的碳阳极在环形炉内被烘烤. 在炉内进行一项实验来确定哪些因子影响烘烤后粘在阳极上的粘贴材料的重量. 考虑 6 个变量, 每个有两个水平: $A=$ 树脂 / 细石比 (0.45, 0.55), $B=$ 粘贴材料类型 (1, 2), $C=$ 粘贴材料温度 (室温, 325), $D=$ 导气管位置 (内部, 外部), $E=$ 槽温 (室温, 195), $F=$ 粘贴前的搁置时间 (0, 24 小时). 做一个 2^{6-3} 设计, 在每一设计点做 3 次重复. 粘在阳极的粘贴材料的重量用克为单位测量. 依实验顺序所得的数据如下.

$$abd=(984,826,936);\quad abcdef=(1\ 275,976,1\ 457);\quad be=(1\ 217,1\ 201,890);$$
$$af=(1\ 474,1\ 164,1\ 541);\quad def=(1\ 320,1\ 156,913);\quad cd=(765,705,821);$$

$$ace=(1\ 338,1\ 254,1\ 294);\quad bcf=(1\ 325,1\ 299,1\ 253)$$

我们希望粘附的粘贴材料重量最小.

(a) 证明此 8 次试验对应于一个 2_{III}^{6-3} 设计, 别名结构是什么?

(b) 用平均重量作为一个响应, 哪些因子是有影响的?

(c) 用重量的极差作为一个响应, 哪些因子是有影响的?

(d) 你会向工程师建议些什么?

8.25 一半导体制造厂进行有 16 个试验的实验, 用来研究 6 个因子对所生产的基层设备弯曲度的效应. 这 6 个变量及其水平如下:

试验	层压温度	层压时间	层压压力	燃烧温度	燃烧周期	燃烧露点
1	55	10	5	1 580	17.5	20
2	75	10	5	1 580	29	26
3	55	25	5	1 580	29	20
4	75	25	5	1 580	17.5	26
5	55	10	10	1 580	29	26
6	75	10	10	1 580	17.5	20
7	55	25	10	1 580	17.5	26
8	75	25	10	1 580	29	20
9	55	10	5	1 620	17.5	26
10	75	10	5	1 620	29	20
11	55	25	5	1 620	29	26
12	75	25	5	1 620	17.5	20
13	55	10	10	1 620	29	20
14	75	10	10	1 620	17.5	26
15	55	25	10	1 620	17.5	20
16	75	25	10	1 620	29	26

每个试验重复做 4 次, 测量基层的弯曲度. 数据如下:

(a) 本实验用了哪一类设计?

(b) 此设计的别名关系是什么?

(c) 哪些过程变量影响平均弯曲度?

(d) 哪些过程变量影响弯曲度的变异性?

(e) 如果尽可能减小弯曲度是重要的, 你会提出什么建议?

试验	弯曲度				总和	均值	标准差
	1	2	3	4	(10^{-4}in/in)	(10^{-4}in/in)	
1	0.016 7	0.012 8	0.014 9	0.018 5	629	157.25	24.418
2	0.006 2	0.006 6	0.004 4	0.002 0	192	48.00	20.976
3	0.004 1	0.004 3	0.004 2	0.005 0	176	44.00	4.083
4	0.007 3	0.008 1	0.003 9	0.003 0	223	55.75	25.025

(续)

试验	弯曲度				总和	均值	标准差
	1	2	3	4	$(10^{-4}$in/in)	$(10^{-4}$in/in)	
5	0.004 7	0.004 7	0.004 0	0.008 9	223	55.75	22.410
6	0.021 9	0.025 8	0.014 7	0.029 6	920	230.00	63.639
7	0.012 1	0.009 0	0.009 2	0.008 6	389	97.25	16.029
8	0.025 5	0.025 0	0.022 6	0.016 9	900	225.00	39.42
9	0.003 2	0.002 3	0.007 7	0.006 9	201	50.25	26.725
10	0.007 8	0.015 8	0.006 0	0.004 5	341	85.25	50.341
11	0.004 3	0.002 7	0.002 8	0.002 8	126	31.50	7.681
12	0.018 6	0.013 7	0.015 8	0.015 9	640	160.00	20.083
13	0.011 0	0.008 6	0.010 1	0.015 8	455	113.75	31.12
14	0.006 5	0.010 9	0.012 6	0.007 1	371	92.75	29.51
15	0.015 5	0.015 8	0.014 5	0.014 5	603	150.75	6.75
16	0.009 3	0.012 4	0.011 0	0.013 3	460	115.00	17.45

8.26 用匀胶机把光刻胶涂在裸硅片上. 这是在早期的半导体制造过程中常用的工艺, 平均涂层厚度和涂层厚度的变异性对下一道生产工序有重要的影响. 该实验有 6 个变量, 这些变量和及其高低水平如下:

因　子	低水平	高水平
最终自旋速度	7 350 转/分	6 650 转/分
加速度	5	20
所用保护材料的体积	3 cc	5 cc
自旋时间	14 s	6 s
保护膜材料的批次	批 1	批 2
排气压强	不盖	盖上

实验者决定用 2^{6-1} 设计, 并在每一片硅片上读取 3 个涂层厚度数据, 这些数据如表 8.37 所示:

(a) 证实这是一个 2^{6-1} 设计, 讨论此设计的别名关系.

(b) 是什么因子影响平均涂层厚度?

(c) 因为使用保护材料的体积对平均厚度的影响很小, 这对生产工程师来说是否有重要的实践意义?

(d) 将该设计投影到仅关于显著性因子的较小设计中, 用图形显示这些结果, 这有助于说明问题吗?

(e) 用涂层厚度的极差作为响应变量. 是否显示出有因子影响涂层厚度的变异性?

(f) 你给工程师推荐什么条件来进行生产?

表 8.37　思考题 8.26 的数据

A	B	C	D	E	F		涂层厚度				
试验	量	批	时间	速度	加速度	盖	左	中间	右	平均	极差
1	5	批 2	14	7 350	5	不盖	4 531	4 531	4 515	4 525.7	16
2	5	批 1	6	7 350	5	不盖	4 446	4 464	4 428	4 446	36
3	3	批 1	6	6 650	5	不盖	4 452	4 490	4 452	4 464.7	38
4	3	批 2	14	7 350	20	不盖	4 316	4 328	4 308	4 317.3	20
5	3	批 1	14	7 350	5	不盖	4 307	4 295	4 289	4 297	18
6	5	批 1	6	6 650	20	不盖	4 470	4 492	4 495	4 485.7	25
7	3	批 1	6	7 350	5	盖上	4 496	4 502	4 482	4 493.3	20
8	5	批 2	14	6 650	20	不盖	4 542	4 547	4 538	4 542.3	9
9	5	批 1	14	6 650	5	不盖	4 621	4 643	4 613	4 625.7	30
10	3	批 1	14	6 650	5	盖上	4 653	4 670	4 645	4 656	25

(续)

A 试验	B 量	C 批	D 时间	E 速度	F 加速度	盖	涂层厚度 左	中间	右	平均	极差
11	3	批 2	14	6 650	20	盖上	4 480	4 486	4 470	4 478.7	16
12	3	批 1	6	7 350	20	不盖	4 221	4 233	4 217	4 223.7	16
13	5	批 1	6	6 650	5	盖上	4 620	4 641	4 619	4 626.7	22
14	3	批 1	6	6 650	20	盖上	4 455	4 480	4 466	4 467	25
15	5	批 2	14	7 350	20	盖上	4 255	4 288	4 243	4 262	45
16	5	批 2	6	7 350	5	盖上	4 490	4 534	4 523	4 515.7	44
17	3	批 2	14	7 350	5	盖上	4 514	4 551	4 540	4 535	37
18	3	批 1	14	6 650	20	不盖	4 494	4 503	4 496	4 497.7	9
19	5	批 2	6	7 350	20	不盖	4 293	4 306	4 302	4 300.3	13
20	3	批 2	6	7 350	5	不盖	4 534	4 545	4 512	4 530.3	33
21	5	批 1	14	6 650	20	盖上	4 460	4 457	4 436	4 451	24
22	3	批 2	6	6 650	5	盖上	4 650	4 688	4 656	4 664.7	38
23	5	批 1	14	7 350	20	不盖	4 231	4 244	4 230	4 235	14
24	3	批 2	6	7 350	20	盖上	4 225	4 228	4 208	4 220.3	20
25	5	批 1	14	7 350	5	盖上	4 381	4 391	4 376	4 382.7	15
26	3	批 2	6	6 650	20	不盖	4 533	4 521	4 511	4 521.7	22
27	3	批 1	14	7 350	20	盖上	4 194	4 230	4 172	4 198.7	58
28	5	批 2	6	6 650	5	不盖	4 666	4 695	4 672	4 677.7	29
29	5	批 1	6	7 350	20	盖上	4 180	4 213	4 197	4 196.7	33
30	5	批 2	6	6 650	20	盖上	4 465	4 496	4 463	4 474.7	33
31	5	批 2	14	6 650	5	盖上	4 653	4 685	4 665	4 667.7	32
32	3	批 2	14	6 650	5	不盖	4 683	4 712	4 677	4 690.7	35

8.27 作者的两个朋友 Harry 和 Judy Peterson-Nedry 在俄勒冈的 Newberg 有一座葡萄园和一个酿酒厂. 他们种植了几种不同的葡萄并酿造葡萄酒. 他们在酿酒生产中将析因设计用于生产工艺和产品的改进. 本问题描述了对他们的 1985 年黑比诺葡萄所做的实验. 在这个实验中最初用到了 8 个变量, 变量如下:

变　量	低水平 (−)	高水平 (+)
A= 黑比诺无性系	玻玛 (Pommard)	威登斯维尔 (Wadenswil)
B= 橡木类型	阿里耶型	特朗赛型
C= 桶龄	陈	新
D= 触媒剂: 酵母片/皮酒囊	香槟酒类	蒙特拉酒类
E= 梗	无	全部
F= 桶的烘烤	轻度	中度
G= 整串	无	10%
H= 发酵温度	低 (最高 75°F)	高 (最高 92°F)

他们决定用 16 次试验的 2_{IV}^{8-1} 设计. 葡萄酒由一个专家组在 1986 年 3 月 8 号进行品尝检测. 每位专家将所品尝的 16 个葡萄酒样本排序, 序号为 1 的酒是最好的. 此设计和品尝专家小组的结论见表 8.38.

(a) Harry 和 Judy 选用的设计的别名关系是什么?

(b) 用平均序号 ($\bar{y}$) 作为响应变量, 分析这些数据并得出的结果. 你会发现, 检查效应估计量的正态概率图是有帮助的.

(c) 用序号的标准差 (或某一恰当的变换, 例如 $\ln s$) 作为响应变量. 关于 8 个变量对葡萄酒质量的变异性的效应, 可以得到什么结论?

(d) 看过结果后, 他们认为专家组中有一成员 (DCM) 对啤酒懂得的要比葡萄酒多, 所以他们决定去掉他的评定结果. 这样做会对 (b) 和 (c) 所得的结果与结论产生什么影响?

(e) 假设在实验刚开始前, Harry 和 Judy 发现在法国订购的用来做实验的 8 只新桶没有按时到达, 所有 16 个试验都只好用旧桶做实验, 如果他们就从设计中去掉列 c, 这一点会对设计的别名关系有怎样影响? 他们需要从头开始并构造新的设计吗?

(f) Harry 和 Judy 根据经验知道, 某些处理组合不可能产生好的结果. 例如, 让 8 个变量都处于高水平来做实验时, 一般会得到低度的葡萄酒, 这一点在 1986 年 3 月 8 日的品尝实验中得到证实. 他们想用这相同的 8 个变量对 1986 的黑比诺葡萄做了一个新的设计, 但是他们不想要 8 个因子都处于高水平的那个试验. 你会建议用什么设计呢?

表 8.38　葡萄酒品尝实验的设计与数据

试验	变量								专家小组排序					统计指标	
	A	B	C	D	E	F	G	H	HPN	JPN	CAL	DCM	RGB	$\bar{y}$	s
1	−	−	−	−	−	−	−	−	12	6	13	10	7	9.6	3.05
2	+	−	−	−	−	+	+	+	10	7	14	14	9	10.8	3.11
3	−	+	−	−	+	−	+	+	14	13	10	11	15	12.6	2.07
4	+	+	−	−	+	+	−	−	9	9	7	9	12	9.2	1.79
5	−	−	+	−	+	+	+	−	8	8	11	8	10	9.0	1.41
6	+	−	+	−	+	−	−	+	16	12	15	16	16	15.0	1.73
7	−	+	+	−	−	+	−	+	6	5	6	5	3	5.0	1.22
8	+	+	+	−	−	−	+	−	15	16	16	15	14	15.2	0.84
9	−	−	−	+	+	+	−	+	1	2	3	3	2	2.2	0.84
10	+	−	−	+	+	−	+	−	7	11	4	7	6	7.0	2.55
11	−	+	−	+	−	+	+	−	13	3	8	12	8	8.8	3.96
12	+	+	−	+	−	−	−	+	3	1	5	1	4	2.8	1.79
13	−	−	+	+	−	−	+	+	2	10	2	4	5	4.6	3.29
14	+	−	+	+	−	+	−	−	4	4	1	2	1	2.4	1.52
15	−	+	+	+	+	−	−	−	5	15	9	6	11	9.2	4.02
16	+	+	+	+	+	+	+	+	11	14	12	13	13	12.6	1.14

8.28 在 *Quality Engineering* 杂志的一篇论文 ("An Application of Fractional Factorial Experimental Designs", 1988, Vol.1, pp. 19~23) 中, M. B. Kilgo 描述了一个实验, 该实验用于确定 CO_2 压强 (A)、CO_2 温度 (B)、花生的湿度 (C)、CO_2 的流速 (D) 和花生的粒度 (E) 对每批花生中油的总产率 (y) 的影响. 她使用的因子水平列在表 8.39 中. 下面是她做的 16 个试验的分式析因实验.

(a) 所用是什么类型的设计? 确定该设计的定义关系和别名关系.

(b) 估计因子的效应, 并用正态概率图初步识别出重要因子.

(c) 用恰当的统计分析方法, 对 (b) 中所找出的对花生油总产率有显著性影响的因子, 并进行假设检验.

(d) 以识别出的重要因子为回归项, 拟合一个可以预测花生油总产率的模型.

(e) 分析此实验的残差, 并评价模型的适合性.

A	B	C	D	E	y	A	B	C	D	E	y
415	25	5	40	1.28	63	415	25	5	60	4.05	23
550	25	5	40	4.05	21	550	25	5	60	1.28	74
415	95	5	40	4.05	36	415	95	5	60	1.28	80
550	95	5	40	1.28	99	550	95	5	60	4.05	33
415	25	15	40	4.05	24	415	25	15	60	1.28	63
550	25	15	40	1.28	66	550	25	15	60	4.05	21
415	95	15	40	1.28	71	415	95	15	60	4.05	44
550	95	15	40	4.05	54	550	95	15	60	1.28	96

表 8.39　思考题 8.28 的实验的因子水平

	A	B	C	D	E
规范水平	压强 (10^5 帕)	温度 (℃)	湿度 (重量百分比)	流速 (升/分钟)	粒度 (mm)
−1	415	25	5	40	1.28
1	550	95	15	60	4.05

8.29 福特汽车公司的埃塞克斯铝厂的工程师进行了一个关于内燃机排气管的砂型铸造实验，实验采用 10 个因子 16 次试验的分式析因设计. D. Becknell 的论文 “Evaporative Cast Process 3.0 Liter Intake Manifold Poor Sandfill Study” (*Fourth Symposium on Taguchi Methods, American Supplier Institue, Dearborn*, MI, 1986, pp. 120~130) 描述了该实验. 实验目的就是确定 10 个因子中哪些因子对铸件的次品率有影响. 这个设计和各个试验观测到的合格铸件比例 $\hat{p}$ 列在表 8.40 中. 这是一个分辨度为III的分式设计，生成元为：$E=CD$, $F=BD$, $G=BC$, $H=AC$, $J=AB$, $K=ABC$. 假设在设计的每个试验中生产的铸件数目都为 1 000.

表 8.40　砂型铸造实验

试验	A	B	C	D	E	F	G	H	J	K	$\hat{p}$	arcsin $\sqrt{\hat{p}}$	F&T 的改进
1	−	−	−	−	+	+	+	+	+	−	0.958	1.364	1.363
2	+	−	−	−	+	+	+	−	−	+	1.000	1.571	1.555
3	−	+	−	−	+	−	−	+	−	+	0.977	1.419	1.417
4	+	+	−	−	+	−	−	−	+	−	0.775	1.077	1.076
5	−	−	+	−	−	+	−	−	+	+	0.958	1.364	1.363
6	+	−	+	−	−	+	−	+	−	−	0.958	1.364	1.363
7	−	+	+	−	−	−	+	−	−	−	0.813	1.124	1.123
8	+	+	+	−	−	−	+	+	+	+	0.906	1.259	1.259
9	−	−	−	+	−	−	+	+	+	−	0.679	0.969	0.968
10	+	−	−	+	−	−	+	−	−	+	0.781	1.081	1.083
11	−	+	−	+	−	+	−	+	−	+	1.000	1.571	1.556
12	+	+	−	+	−	+	−	−	+	−	0.896	1.241	1.242
13	−	−	+	+	+	−	−	−	+	+	0.958	1.364	1.363
14	+	−	+	+	+	−	−	+	−	−	0.818	1.130	1.130
15	−	+	+	+	+	+	+	−	−	−	0.841	1.161	1.160
16	+	+	+	+	+	+	+	+	+	+	0.955	1.357	1.356

(a) 找出这个设计的定义关系和别名关系.

(b) 估计因子的效应，并用正态概率图初步识别出重要因子.

(c) 用上面 (b) 中所识别出的因子拟合恰当的模型.

(d) 画出模型关于预测铸件合格率的残差图. 并画出残差的正态概率图. 评价模型适合性.

(e) 在 (d) 中你应该注意到响应的方差并非常数这一信息. (考虑到响应是比例值, 你也应该能估计到这点.) 先前的表格也列出了 $\hat{p}$ 的一个变换 ——arcsin 的平方根变换, 它是一个广泛用于比例数据的稳定方差变换 (参见第 3 章中关于稳定方差的讨论). 用这种变换后的响应, 重做上述的 (a) 到 (d), 并评论你得到的结果. 特别是, 残差图被改进了吗?

(f) 有一种修改的反正弦平方根变换, 它是由 Freeman 和 Tukey 提出的 ("Transformations Related to the Angular and the Square Root." *Annals of Mathematical Statistics*, Vol. 21, 1950, pp. 607~611), 它改进了关于尾部的性质. F&T 的修改为:

$$\left[\arcsin\sqrt{n\hat{p}/(n+1)} + \arcsin\sqrt{(n\hat{p}+1)/(n+1)}\right]/2$$

用这个变换重做 (a) 到 (d), 并评论你得到的结果. (要知关于此实验的一个有趣的讨论与分析, 参见 S. Bisgaard 和 H. T. Fuller 的 "Analysis of Factorial Experiments with Defects or Defectives as the Response", *Quality Engineering*, Vol. 7, 1994-1995, pp. 429~443.)

8.30 克莱斯勒汽车公司的工程师进行了一个有 9 个因子 16 次试验的分式析因实验, P. I. Hsieh 和 D. E. Goodwin 的文章 "Sheet Molded Compound Process Improvement" (*Fourth Symposium on Taguchi Methods*, American Supplier Institue, Dearborn, MI, 1986, pp. 13~21) 描述了这个实验. 实验目的就是减少可开式模压护栅板的末道漆的缺陷数. 这个设计, 以及各次试验中观测到的缺陷数 c, 列在表 8.41 中. 这是一个分辨度为III的分式设计, 生成元为: $E = BD$, $F = BCD$, $G = AC$, $H = ACD$, $J = AB$.

表 8.41　护栅缺陷实验

试验	A	B	C	D	E	F	G	H	J	c	$\sqrt{c}$	F&T 的改进
1	−	−	−	−	+	−	+	−	+	56	7.48	7.52
2	+	−	−	−	+	−	−	+	−	17	4.12	4.18
3	−	+	−	−	−	+	+	−	−	2	1.41	1.57
4	+	+	−	−	−	+	−	+	+	4	2.00	2.12
5	−	−	+	−	+	+	−	+	+	3	1.73	1.87
6	+	−	+	−	+	+	+	−	−	4	2.00	2.12
7	−	+	+	−	−	−	−	+	−	50	7.07	7.12
8	+	+	+	−	−	−	+	−	+	2	1.41	1.57
9	−	−	−	+	−	+	+	+	+	1	1.00	1.21
10	+	−	−	+	−	+	−	−	−	0	0.00	0.50
11	−	+	−	+	+	−	+	+	−	3	1.73	1.87
12	+	+	−	+	+	−	−	−	+	12	3.46	3.54
13	−	−	+	+	−	−	−	−	+	3	1.73	1.87
14	+	−	+	+	−	−	+	+	−	4	2.00	2.12
15	−	+	+	+	+	+	−	−	−	0	0.00	0.50
16	+	+	+	+	+	+	+	+	+	0	0.00	0.50

(a) 找出这个设计的定义关系和别名关系.

(b) 估计因子的效应, 并用正态概率图初步识别出重要因子.

(c) 用上面 (b) 中所识别出的因子拟合恰当的模型.

(d) 画出关于模型预测缺陷数的残差图. 并画出残差的正态概率图. 评价模型的适合性.

(e) 在 (d) 中你应该注意到响应的方差并非常数这一信息. (考虑到响应是计数型变量, 你也应该能估计到这点.) 先前的表格也列出了 c 的一个变换 —— 平方根变换, 它被广泛用于计数型数据的方差稳定变换 (参见第 3 章中关于稳定方差的讨论). 用这变换后的响应, 重做上述的 (a) 到 (d), 并评论你得到的结果. 特别是, 残差图被改进了吗?

(f) 有一种修改的平方根变换, 它是由 Freeman 和 Tukey 提出的 ("Transformations Related to the Angular and the Square Root," *Annals of Mathematical Statistics*, Vol. 21, 1950, pp.

607~611), 它改进了原先的性能. F&T 对平方根变换的修改为:

$$[\sqrt{c}+\sqrt{c+1}]/2$$

用这个变换重做 (a) 到 (d), 并评论你的结果. (要知关于此实验的一个有趣的讨论与分析, 请参见 S. Bisgaard 和 H. T. Fuller 的 "Analysis of Factorial Experiments with Defects or Defectives as the Response", *Quality Engineering*, Vol. 7, 1994~1995, pp. 429~443.)

8.31 为了研究 6 个因子对于晶体管增益的影响, 一个半导体工厂进行了一个实验. 该实验选择的设计是下面的 2_{IV}^{6-2} 设计:

标准序号	试验序号	A	B	C	D	E	F	增益
1	2	−	−	−	−	−	−	1 455
2	8	+	−	−	−	+	−	1 511
3	5	−	+	−	−	+	+	1 487
4	9	+	+	−	−	−	+	1 596
5	3	−	−	+	−	+	+	1 430
6	14	+	−	+	−	−	+	1 481
7	11	−	+	+	−	−	−	1 458
8	10	+	+	+	−	+	−	1 549
9	15	−	−	−	+	−	+	1 454
10	13	+	−	−	+	+	+	1 517
11	1	−	+	−	+	+	−	1 487
12	6	+	+	−	+	−	−	1 596
13	12	−	−	+	+	+	−	1 446
14	4	+	−	+	+	−	−	1 473
15	7	−	+	+	+	−	+	1 461
16	16	+	+	+	+	+	+	1 563

(a) 用效应的正态概率图识别显著的因子.

(b) 对 (a) 中识别出的显著因子作合适的统计检验.

(c) 分析残差并评论你的发现.

(d) 你能找出一组使产品增益在 1 500±25 内的运行条件吗?

8.32 热处理常用于碳化金属零件, 如齿轮. 碳化层的厚度是该工序的关键输出变量, 它通常对齿轮的齿距 (齿轮的齿尖) 进行碳分析来测量. 此 2_{IV}^{6-2} 设计要研究 6 个因子: A= 炉温, B= 循环时间, C= 碳浓度, D= 碳化持续时间, E= 扩散循环的碳浓度, F= 扩散循环的持续时间. 实验结果如表中所列.

标准序号	试验序号	A	B	C	D	E	F	齿距
1	5	−	−	−	−	−	−	74
2	7	+	−	−	−	+	−	190
3	8	−	+	−	−	+	+	133
4	2	+	+	−	−	−	+	127
5	10	−	−	+	−	+	+	115
6	12	+	−	+	−	−	+	101
7	16	−	+	+	−	−	−	54
8	1	+	+	+	−	+	−	144
9	6	−	−	−	+	−	+	121
10	9	+	−	−	+	+	+	188
11	14	−	+	−	+	+	−	135
12	13	+	+	−	+	−	−	170
13	11	−	−	+	+	+	−	126
14	3	+	−	+	+	−	−	175
15	15	−	+	+	+	−	+	126
16	4	+	+	+	+	+	+	193

(a) 估计因子效应并画出它们的正态概率图, 选择一个初始模型.

(b) 对这个模型进行恰当的统计检验.

(c) 分析残差, 并评价模型的适合性.

(d) 解释该实验的结果, 假定需要的碳化层的厚度在 140 至 160 之间.

8.33 L. B. Hare 的一篇文章 ("In the Soup: A Case Study to Identify Contributors to Filling Variability", *Journal of Quality Technology*, Vol. 20, pp. 36~43) 里描述了一个用于研究干汤料包的装填重量变异性的析因实验. 因子是: A= 蔬菜油的加料口个数 (1, 2), B= 混合物的环境温度 (冷, 微热), C= 混合时间 (60 秒, 80 秒), D= 批的重量 (1 500lb, 2 000 lb), E= 混合与包装的间隔天数 (1, 7). 设计的每个试验中, 8 小时内被抽取的汤料包数量在 125 到 150 之间, 汤料包重量的标准差作为响应变量. 设计与所得的结果见下表.

标准序号	A 混料口	B 温度	C 时间	D 批的重量	E 搁置时间	y 标准差
1	−	−	−	−	−	1.13
2	+	−	−	−	+	1.25
3	−	+	−	−	+	0.97
4	+	+	−	−	−	1.7
5	−	−	+	−	+	1.47
6	+	−	+	−	−	1.28
7	−	+	+	−	−	1.18
8	+	+	+	−	+	0.98
9	−	−	−	+	+	0.78
10	+	−	−	+	−	1.36
11	−	+	−	+	−	1.85
12	+	+	−	+	+	0.62
13	−	−	+	+	−	1.09
14	+	−	+	+	+	1.1
15	−	+	+	+	+	0.76
16	+	+	+	+	−	2.1

(a) 这个估计的生成元是什么?

(b) 这个设计的分辨度是什么?

(c) 估计各因子效应, 哪些效应是大的?

(d) 残差分析是否表明了采用的假定有问题?

(e) 对这个装填工序能得到什么结论?

8.34 考虑 2_{IV}^{6-2} 设计.

(a) 假设通过改变列 B 的符号取代改变列 A 的符号, 折叠了此设计, 则该组合的设计估计的效应会有什么改变?

(b) 假设通过改变列 E 的符号取代改变列 A 的符号, 折叠了此设计, 则该组合的设计估计的效应会有什么改变?

8.35 考虑 2_{IV}^{7-3} 设计. 若通过改变列 A 的符号折叠此设计, 确定该组合的设计的别名关系.

8.36 考虑思考题 8.34 中的 2_{IV}^{7-3} 设计.

(a) 若通过改变列 B 的符号折叠此设计, 确定该组合的设计的别名关系.

(b) 比较这个组合的设计与问题 8.35 中得到的组合的设计的别名, 改变不同列的符号会导致什么不同的结果?

8.37 考虑 2_{IV}^{7-3} 设计.

(a) 假设此设计的部分折叠是通过列 A (只取正号) 得到, 决定组合的设计的别名关系.

(b) 取列 A 的负号来定义部分折叠, 重做上述 (a). 用不同符号定义部分折叠是否有所不同?

8.38 考虑 2_{IV}^{6-2} 的一个部分折叠. 假设这个设计的部分折叠是取列 A 的相反符号得到, 但保留的 8 个试验是列 C 取正号的那些试验, 确定该组合设计别名关系.

8.39 考虑 2_{III}^{7-4} 的部分折叠. 假设这个设计的部分折叠是由列 A (只取正号) 构造的. 确定该组合设计的别名关系.

8.40 考虑 2_{III}^{5-2} 的部分折叠. 假设这个设计的部分折叠是由列 A (只取正号) 构造的. 确定该组合设计的别名关系.

8.41 重新考虑例 8.1 的 2^{4-1} 设计. 显著因子为: $A, C, D, AC + BD, AD + BC$. 找出能使交互作用 AC, BD, AD, BC 可被估计的部分折叠设计.

第 9 章　三水平和混合水平析因设计与分式析因设计

本章纲要

9.1　3^k 析因设计
　9.1.1　3^k 设计的记号和引进动机
　9.1.2　3^2 设计
　9.1.3　3^3 设计
　9.1.4　一般的 3^k 设计
9.2　3^k 析因设计的混区设计
　9.2.1　三区组的 3^k 析因设计
　9.2.2　九区组的 3^k 析因设计
　9.2.3　3^p 个区组的 3^k 析因设计
9.3　3^k 析因设计的分式重复
　9.3.1　3^k 析因设计的 $\frac{1}{3}$ 分式设计
　9.3.2　其他的 3^{k-p} 分式析因设计
9.4　混合水平的析因设计
　9.4.1　二水平和三水平的因子
　9.4.2　二水平和四水平的因子
第 9 章补充材料
S9.1　3^k 设计的 Yates 算法
S9.2　三水平和混合水平设计中的别名

第 6 章至第 8 章讨论的二水平析因设计和分式析因设计系列广泛应用于工业研究与开发. 有时会使用这些设计的某些推广和变形, 例如所有因子都为三水平的设计. 本章将讨论这些 3^k 设计. 我们还将考虑某些因子为二水平, 另一些因子为三水平或四水平的设计.

9.1　3^k 析因设计

9.1.1　3^k 设计的记号和引进动机

现在讨论 3^k 析因设计, 也就是, 有 k 个因子、每个因子有 3 个水平的析因设计. 因子和交互作用将用大写字母表示. 我们把因子的 3 个水平看作低、中、高. 这些因子水平可用几种不同的记号表达, 其中之一是以数字 0(低)、1(中)、2(高) 表示因子各水平. 3^k 设计的每个处理组合用 k 个数字表示, 其中第 1 个数字表示因子 A 的水平, 第 2 个数字表示因子 B 的水平, $\cdots$, 第 k 个数字表示因子 K 的水平. 例如, 在 3^2 设计中, 00 表示对应于 A 和 B 都处于低水平的处理组合, 01 表示对应于 A 处于低水平、B 处于中水平的处理组合. 图 9.1 和图 9.2 用这种记号分别给出了 3^2 和 3^3 设计的几何表达.

这一记号系统也可用于前面介绍的 2^k 设计, 只要分别用 0 和 1 代替 −1 和 +1 即可. 在 2^k 设计中, 我们宁愿用 ±1 记号, 因为它使设计的几何观点变得更为方便, 而且它可以直接应用于回归模型、区组化以及分式析因设计的建构.

在 3^k 设计系统中, 对于**定量**因子, 我们常把它们的低、中、高水平分别记为 $-1, 0, +1$. 这便于**拟合响应关于因子水平的回归模型** (regression model). 例如, 考虑图 9.1 中的 3^2 设计, 设 x_1 表示因子 A, x_2 表示因子 B. 响应 y 关于 x_1 和 x_2 的回归模型可记为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon \tag{9.1}$$

注意, 增加因子的第 3 个水平就容许响应和设计因子之间的关系用二次函数模拟.

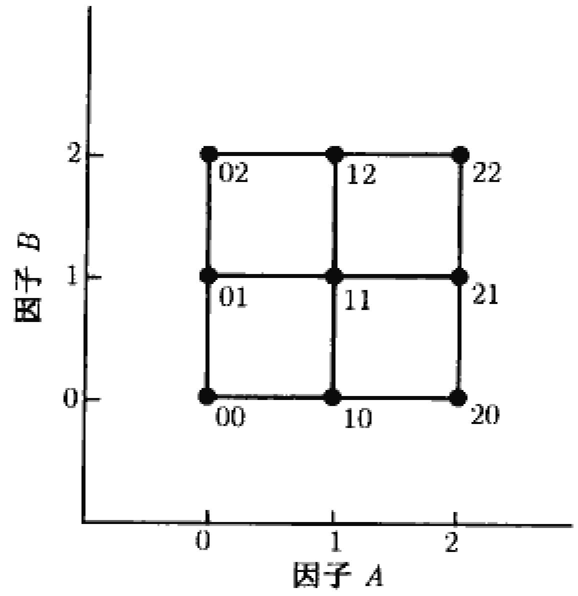


图 9.1 3^2 设计的处理组合

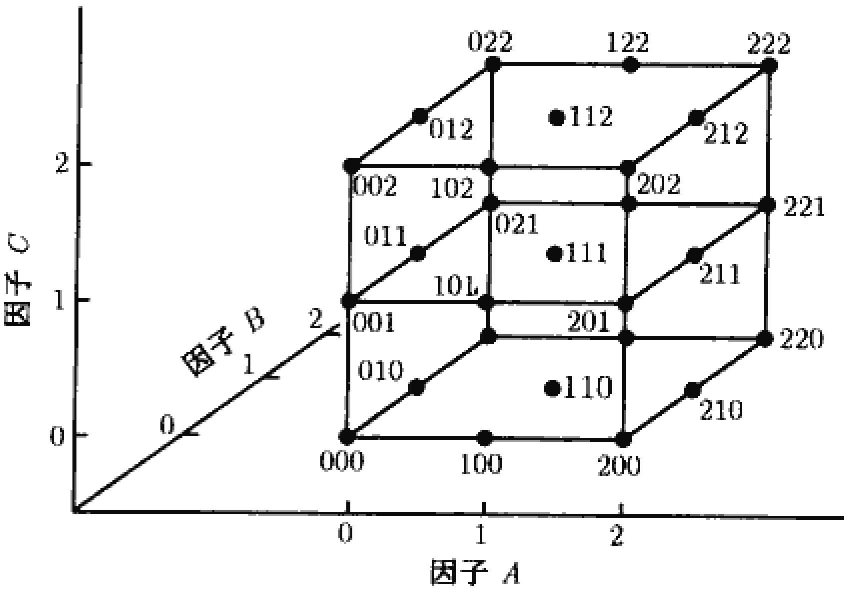


图 9.2 3^3 设计的处理组合

3^k 设计的确是关心响应函数弯曲性的实验者的一个可能选择. 不过, 有两点必须考虑:

(1) 3^k 设计不是模拟二次关系最有效的方法. 第 11 章讨论的**响应曲面设计**是更好的方法.

(2) 如第 6 章所讨论的, 添加中心点的 2^k 设计是得到弯曲性特征的极好方法. 它容许人们把设计的大小和复杂性保持在低的程度上, 并同时得出一些针对弯曲性的保护措施. 因此, 如果弯曲性很重要, 就可以如图 6.37 所示, 在二水平设计的基础上增加若干**轴点上的试验** (axial run) 来获得**中心复合设计**. 这种实验的**序贯方法**比带有定量因子的 3^k 设计的效率高得多.

9.1.2 3^2 设计

3^k 系统中最简单的设计是 3^2 设计, 它有两个因子, 每个因子有 3 个水平. 此设计的处理组合见图 9.1. 因为有 $3^2=9$ 个处理组合, 所以这些处理组合间有 8 个自由度. A 和 B 的主效应各有两个自由度, AB 交互作用有 4 个自由度. 如果有 n 次重复, 则有 3^2n-1 个总的自由度和 $3^2(n-1)$ 个误差自由度.

A, B, AB 的平方和可用第 5 章讨论的关于析因设计的通常方法来计算. 如 (9.1) 式所示, 每个主效应可表达为一个线性分量和一个二次分量, 每一分量都是单自由度的. 当然, 这只在因子是定量的时候才有意义.

二因子交互作用 AB 可用两种方法分解. 第一种方法是把 AB 分为 4 个单自由度的分量 $AB_{L\times L}$, $AB_{L\times Q}$, $AB_{Q\times L}$, $AB_{Q\times Q}$. 这些分量分别可通过拟合项 $\beta_{12}x_1x_2$, $\beta_{122}x_1x_2^2$, $\beta_{112}x_1^2x_2$, $\beta_{1122}x_1^2x_2^2$ 得到, 并已在例 5.5 中作了具体说明. 对于刀具寿命数据, 可以得出 $SS_{AB_{L\times L}}=8.00$, $SS_{AB_{L\times Q}}=42.67$, $SS_{AB_{Q\times L}}=2.67$, $SS_{AB_{Q\times Q}}=8.00$. 因为这是 AB 的正交分解, 所以 $SS_{AB}=SS_{AB_{L\times L}}+SS_{AB_{L\times Q}}+SS_{AB_{Q\times L}}+SS_{AB_{Q\times Q}}=61.34$.

第二种方法基于**正交拉丁方**. 考虑例 5.5 中各处理组合下数据的和, 这些和列于图 9.3 中, 就是方形内圈起来的数, 两个因子 A 和 B 分别对应于某一 3×3 拉丁方的行和列. 图 9.3 是两个特定的、叠加了单元和的 3×3 拉丁方.

这两个拉丁方是**正交的**, 也就是说, 如果将一个拉丁方叠在另一拉丁方之上, 则第一个拉丁

方的每一个字母和第二个拉丁方的每一个字母在一起只出现一次. 拉丁方 (a) 的字母总和是 $Q = 18$, $R = -2$, $S = 8$, 其平方和是 $[18^2 + (-2)^2 + 8^2]/(3)(2) - [24^2/(9)(2)] = 33.34$, 有两个自由度. 同理, 拉丁方 (b) 的字母总和是 $Q = 0$, $R = 6$, $S = 18$, 其平方和是 $[0^2 + 6^2 + 18^2]/(3)(2) - [24^2/(9)(2)] = 28.00$, 有两个自由度. 这两个分量的和是

$$33.34 + 28.00 = 61.34 = SS_{AB}$$

有 $2+2=4$ 个自由度.

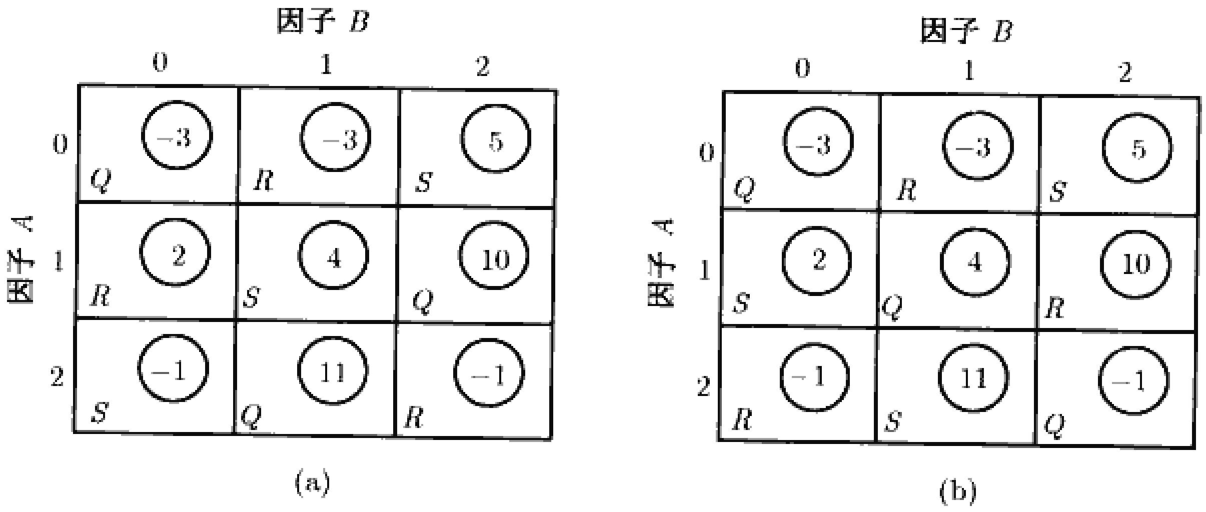


图 9.3　例 5.5 处理组合内的和叠加在两个正交的拉丁方上

一般地, 由拉丁方 (a) 算得的平方和叫做**交互作用的 AB 分量**, 由拉丁方 (b) 算得的平方和叫做**交互作用的 AB^2 分量**. 分量 AB 和 AB^2 各有两个自由度. 之所以使用这一专门名词, 是因为, 如果分别用 x_1 和 x_2 记 A 和 B 的水平 (0, 1, 2) , 则会发现单元上的字母服从如下规律:

拉丁方 (a)	拉丁方 (b)
$Q: x_1 + x_2 = 0 \pmod 3$	$Q: x_1 + 2x_2 = 0 \pmod 3$
$R: x_1 + x_2 = 1 \pmod 3$	$S: x_1 + 2x_2 = 1 \pmod 3$
$S: x_1 + x_2 = 2 \pmod 3$	$R: x_1 + 2x_2 = 2 \pmod 3$

例如, 在拉丁方 (b) 中, 中间单元对应于 $x_1 = 1$ 和 $x_2 = 1$, 于是 $x_1 + 2x_2 = 1 + (2)(1) = 3 = 0 \pmod 3$, 从而 Q 占据正中央单元. 当考虑形如 A^pB^q 的表达式时, 为方便起见, 只允许第一个字母的指数是 1. 如果第一个字母的指数不是 1 的话, 则将整个表达式平方, 然后将指数按模数为 3 化简. 例如, A^2B 和 AB^2 相同, 因为

$$A^2B = (A^2B)^2 = A^4B^2 = AB^2$$

AB 交互作用的分量 AB 和分量 AB^2 没有实际意义, 通常也不在方差分析表中显示出来. 不过, 在构造更为复杂的设计时, 把 AB 交互作用分解为两个正交的二自由度的分量是很有用的. 还有, 交互作用的分量 AB 和分量 AB^2 和 $AB_{L\times L}$, $AB_{L\times Q}$, $AB_{Q\times L}$, $AB_{Q\times Q}$ 的平方和之间没有关系.

交互作用的分量 AB 和分量 AB^2 可以用另外的方法计算. 考虑图 9.3 拉丁方的各处理组合的和. 如果把从左上到右下的对角线的数加起来, 就得到和数 $-3+4-1=0, -3+10-1=6, 5+11+2=18$. 这些和数之间的平方和是 $28.00(AB^2)$. 同理, 从右上到左下对角线上的和数

是 $5+4-1=8, -3+2-1=-2, -3+11+10=18$. 这些和数之间的平方和是 33.34($AB$). Yates 称交互作用的这些分量为**交互作用的 I 分量和 J 分量**. 交替使用这两种记号, 也就是,

$$I(AB) = AB^2, \qquad J(AB) = AB$$

9.1.3 3^3 设计

现在假定有 3 个因子 (A, B, C) 要研究, 每个因子有 3 个水平, 安排析因设计实验. 这是一个 3^3 析因设计, 实验的安排和处理组合的记号见图 9.2. 27 个处理组合有 26 个自由度. 每个主效应有 2 个自由度, 每个二因子交互作用有 4 个自由度, 三因子交互作用有 8 个自由度. 如果有 n 次重复, 则有 3^3n-1 个总自由度和 $3^3(n-1)$ 个误差自由度.

平方和可用析因设计的标准方法计算, 此外, 如果因子是定量的, 则主效应可分解为单自由度的线性分量和二次分量. 二因子交互作用可分解为线性 × 线性、线性 × 二次、二次 × 线性以及二次 × 二次的效应. 最后, 三因子交互作用可分解为 8 个单自由度的分量, 对应于线性 × 线性 × 线性、线性 × 线性 × 二次, 等等. 三因子交互作用的这种分解一般说来不大有用.

也可以把二因子交互作用分解为它的 I 分量和 J 分量. 这些分量记为 $AB, AB^2, AC, AC^2, BC, BC^2$, 每个分量有两个自由度. 和 3^2 设计一样, 这些分量没有具体意义.

三因子交互作用 ABC 可分解为 4 个正交的二自由度分量, 通常叫做交互作用的 W, X, Y, Z 分量, 也可以分别记为 ABC 交互作用的 AB^2C^2, AB^2C, ABC^2, ABC 分量. 两种记号交替使用, 也就是,

$$W(ABC) = AB^2C^2, \qquad X(ABC) = AB^2C,$$
$$Y(ABC) = ABC^2, \qquad Z(ABC) = ABC$$

注意第一个字母的指数不能异于 1. 和 I 分量与 J 分量相同, W, X, Y, Z 分量没有实际意义. 不过, 它们可用于构造更为复杂的设计.

例 9.1 一台机器用来把软饮料糖浆灌注在 5 加仑的金属容器内. 感兴趣的变量是由起泡沫引起的糖浆损失量. 设想影响起泡沫的有 3 个因子: 喷嘴设计 (A)、灌注速度 (B) 和操作压强 (C). 选取 3 种喷嘴、3 种灌注速度和 3 种操作压强进行有两次重复的 3^3 析因实验. 规范数据见表 9.1.

表 9.1 例 9.1 的糖浆损失数据 (单位: 立方厘米 −70)

压强 (psi) (C)	喷嘴类型 (A) 1			2			3		
	速度 (转/分)(B)								
	100	120	140	100	120	140	100	120	140
10	−35	−45	−40	17	−65	20	−39	−55	15
	−25	−60	15	24	−58	4	−35	−67	−30
15	110	−10	80	55	−55	110	90	−28	110
	75	30	54	120	−44	44	113	−26	135
20	4	−40	31	−23	−64	−20	−30	−61	54
	5	−30	36	−5	−62	−31	−55	−52	4

糖浆损失数据的方差分析见表 9.2. 用通常的方法计算平方和. 我们看到灌注速度和操作压强在统计意义上是显著的. 所有 3 个二因子交互作用都是显著的. 二因子交互作用的图解分析见图 9.4. 中等水平的速度效果最好, 喷嘴类型 2 和类型 3、低压强 (10 psi) 或高压强 (20 psi) 对减少糖浆损失看来最有效.

例 9.1 说明了三水平设计经常应用的情况：一个或多个因子是定性的，自然地取 3 个水平，其余因子是定量的. 本例中假设只对 3 种喷嘴设计有兴趣. 显然它是具有所需的有 3 个水平的定性因子. 灌注速度和操作压强都是定量因子. 因此，可以在喷嘴因子的每个水平上，分别拟合两个因子 (速度和压强)，如 (9.1) 式所示的二次模型.

表 9.2　糖浆损失数据的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
A, 喷嘴	993.77	2	496.89	1.17	0.325 6
B, 速度	61 190.33	2	30 595.17	71.74	< 0.000 1
C, 压强	69 105.33	2	34 552.67	81.01	< 0.000 1
AB	6 300.90	4	1 575.22	3.69	0.016 0
AC	7 513.90	4	1 878.47	4.40	0.007 2
BC	12 854.34	4	3 213.58	7.53	0.000 3
ABC	4 628.76	8	578.60	1.36	0.258 0
误差	11 515.50	27	426.50		
总和	174 102.83	53			

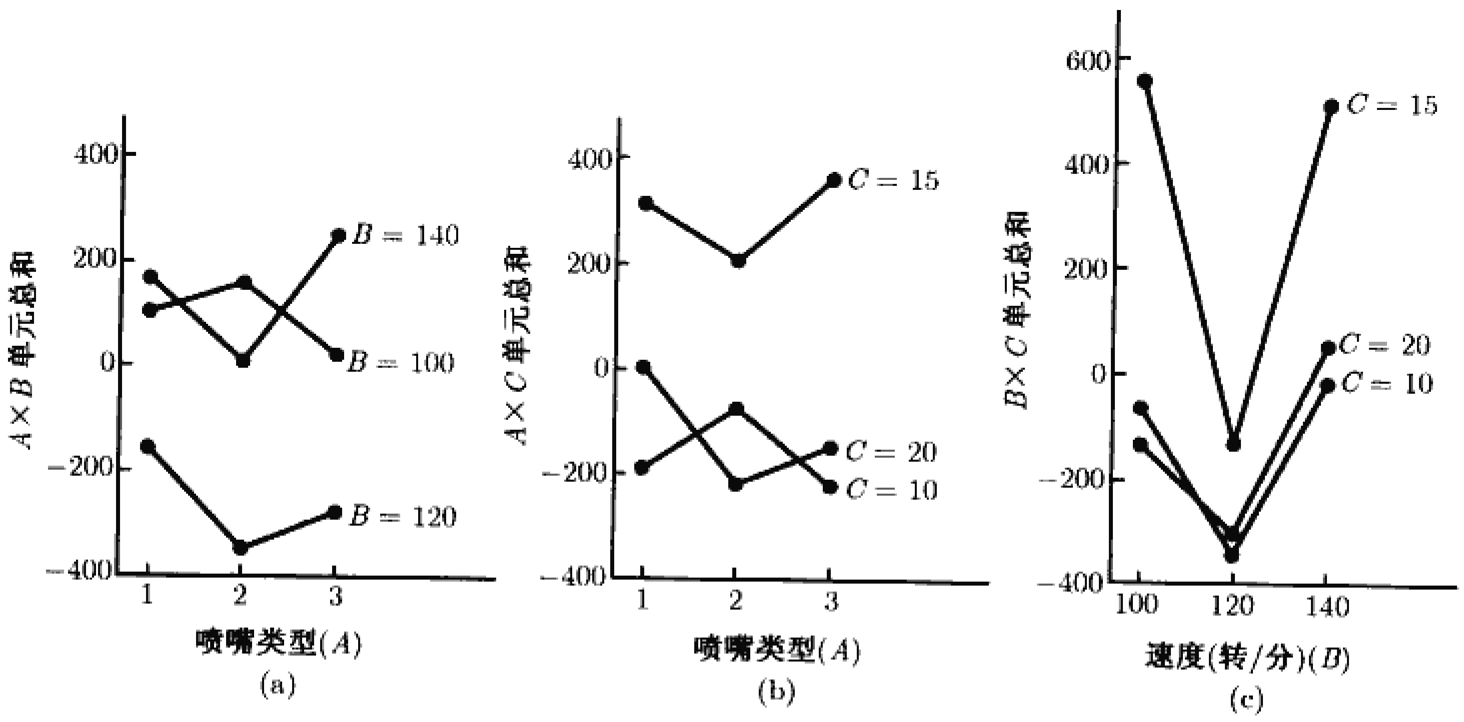


图 9.4　例 9.1 的二因子交互作用

表 9.3 列出了这些二次回归模型. 在这些模型中，用标准线性回归的计算机程序估计 β. (第 10 章将详细讨论最小二乘回归法.) 在这些模型中，变量 x_1 和 x_2 像前面一样被规范为 −1, 0, +1, 可以对压强和速度的自然水平指定下列规范水平：

规范水平	速度 (转/分)	压强 (psi)
−1	100	10
0	120	15
+1	140	20

表 9.3 分别以这些规范水平和速度与压强的自然水平给出了相应的模型.

图 9.5 显示了在每一种喷嘴类型下，作为速度和压强的函数的糖浆损失为常数时的响应曲面等高线图. 这些图揭示了此灌注系统性能的大量的有用信息. 由于实验目标是最小化糖浆损

失, 而所观测到的最小的等高线 (–60) 仅出现在喷嘴类型 3 的等高线图中, 所以可以选用喷嘴类型 3. 应该选接近中间水平 120 转/分的灌注速度, 以及高水平或低水平的压强.

表 9.3 例 9.1 的回归模型

喷嘴类型	x_1= 速度 (S), x_2= 压强 (P) 的规范单位
1	$\hat{y} = 22.1 + 3.5x_1 + 16.3x_2 + 51.7x_1^2 - 71.8x_2^2 + 2.9x_1x_2$
	$\hat{y} = 1\ 217.3 - 31.256S + 86.017P + 0.129\ 17S^2 - 2.873\ 3P_2^2 + 0.028\ 75SP$
2	$\hat{y} = 25.6 - 22.8x_1 - 12.3x_2 + 14.1x_1^2 - 56.9x_2^2 - 0.7x_1x_2$
	$\hat{y} = 180.1 - 9.475S + 66.75P + 0.035S^2 - 2.276\ 7P_2^2 - 0.007\ 5SP$
3	$\hat{y} = 15.1 + 20.3x_1 + 5.9x_2 + 75.8x_1^2 - 94.9x_2^2 + 10.5x_1x_2$
	$\hat{y} = 1\ 940.1 - 46.058S + 102.48P + 0.189\ 58S^2 - 3.796\ 7P_2^2 + 0.105SP$

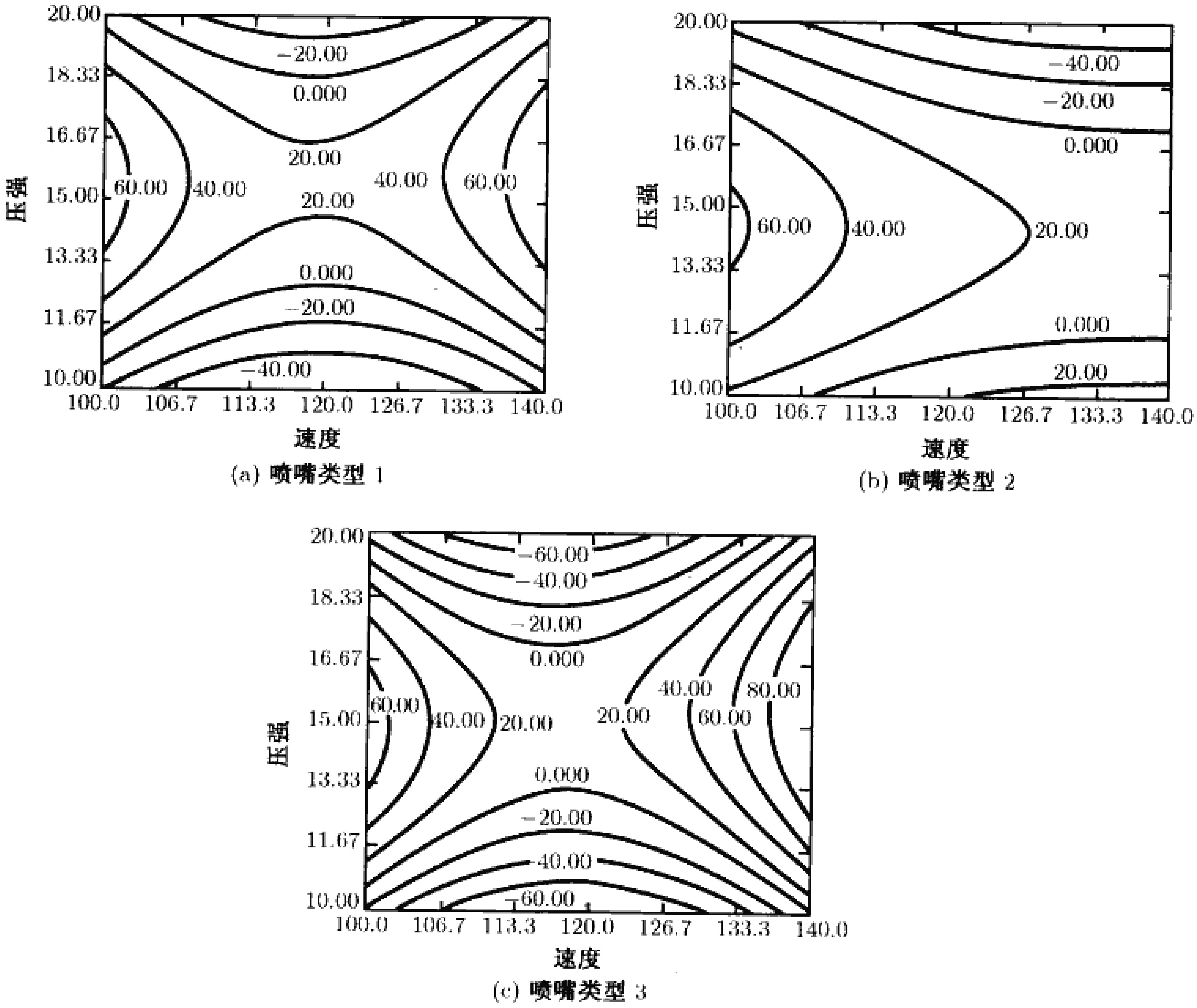


图 9.5 例 9.1 中对喷嘴类型 1, 2, 3, 作为速度和压强函数的糖浆损失 (单位: cc –70) 为常数时的等高线图

对同时具有定量和定性因子的实验构造等高线图时, 通常会发现, 在定性因子的不同水平下, 定量因子的曲面形状有极大的差异. 图 9.5 在某些程度上有醒目的差异, 喷嘴类型 2 的曲面形状明显比喷嘴类型 1 与类型 3 的曲面狭长. 此时, 这意味着在定性因子的不同水平下, 定量因子的

最优操作条件 (和其他重要结论) 有极大的差异.

我们用例 9.1 的数据来具体说明 ABC 交互作用是如何分解为它的 4 个正交的二自由度的分量的. Cochran and Cox (1957) 和 Davies (1956) 描述了一般的方法. 首先, 选取 3 个因子中的任意两个, 比方说 AB, 计算 AB 交互作用在第 3 个因子 C 的各水平处的 I 总和与 J 总和. 这些计算如下:

C	B	A			总和	
		1	2	3	I	J
	100	−60	41	−74	−198	−222
10	120	−105	−123	−122	−106	−79
	140	−25	24	−15	−155	−158
	100	185	175	203	331	238
15	120	20	−99	−54	255	440
	140	134	154	245	377	285
	100	9	−28	−85	−59	−144
20	120	−70	−126	−113	−74	−40
	140	67	−51	58	−206	−155

现在, $I(AB)$ 与 $J(AB)$ 的总和与因子 C 一起以双向表方式列出, 并在这个新显示方式表中计算出 I 和 J 的对角线总和:

C	$I(AB)$			总和		C	$J(AB)$			总和	
				I	J					I	J
10	−198	−106	−155	−149	41	10	−222	−79	−158	63	138
15	331	255	377	212	19	15	238	440	285	62	4
20	−59	−74	−206	102	105	20	−144	−40	−155	40	23

上面算得的 I 与 J 的对角和实际上代表量 $I[I(AB)\times C]=AB^2C^2$, $J[I(AB)\times C]=AB^2C$, $I[J(AB)\times C]=ABC^2$ 以及 $J[J(AB)\times C]=ABC$, 即 ABC 的 W, X, Y, Z 分量. 用通常的方法求平方和, 也就是,

$$I[I(AB)\times C]=AB^2C^2=W(ABC)=\frac{(-149)^2+(212)^2+(102)^2}{18}-\frac{(165)^2}{54}=3\ 804.11$$

$$J[I(AB)\times C]=AB^2C=X(ABC)=\frac{(41)^2+(19)^2+(105)^2}{18}-\frac{(165)^2}{54}=221.77$$

$$I[J(AB)\times C]=ABC^2=Y(ABC)=\frac{(63)^2+(62)^2+(40)^2}{18}-\frac{(165)^2}{54}=18.77$$

$$J[J(AB)\times C]=ABC=Z(ABC)=\frac{(138)^2+(4)^2+(23)^2}{18}-\frac{(165)^2}{54}=584.11$$

虽然这是 SS_{ABC} 的一种正交分解, 但我们再次指出, 习惯上不把它显示在方差分析表中. 在随后各节中, 我们有机会讨论这些分量的一个或多个计算法.

9.1.4　一般的 3^k 设计

在 3^2 设计和 3^3 设计中所用的概念, 可以立即推广到有 k 个因子、每个因子有 3 个水平的

情况, 即 3^k 析因设计中去. 处理组合用通常的记号, 所以 0120 表示 3^k 设计的一个处理组合, 其 A 和 D 处于低水平, B 处于中水平, C 处于高水平. 共有 3^k 个处理组合, 它们之间有 3^k-1 个自由度, 这些处理组合可以确定 k 个二自由度的主效应的平方和, $\binom{k}{2}$ 个四自由度的二因子交互作用的平方和, $\cdots$, 以及一个 2^k 自由度的 k 因子交互作用的平方和. 一般地, h 个因子交互作用有 2^h 个自由度. 如果有 n 次重复, 则有 3^kn-1 个总自由度和 $3^k(n-1)$ 个误差自由度.

用通常对析因设计使用的方法计算效应和交互作用的平方和. 三因子交互作用及更高阶的交互作用一般不再进一步分解. 不过, 任一 h 个因子的交互作用有 2^{h-1} 个正交的二自由度分量. 例如, 四因子的交互作用 $ABCD$, 有 $2^{4-1}=8$ 个正交的二自由度分量, 记为 $ABCD^2$, ABC^2D, AB^2CD, $ABCD$, ABC^2D^2, AB^2C^2D, AB^2CD^2, 以及 $AB^2C^2D^2$. 在写这些分量时, 第一个字母的指数只能是 1. 如果第一个字母的指数不是 1, 则将整个式子平方一下, 然后对指数按模数 3 简化. 为了表明这一点, 请看

$$A^2BCD=(A^2BCD)^2=A^4B^2C^2D^2=AB^2C^2D^2$$

这些交互作用分量没有具体意义, 但在构造更复杂的设计时会有用.

设计的大小随着 k 的增加而迅速增加. 例如, 3^3 设计每一重复有 27 个处理组合, 3^4 设计有 81 个, 3^5 设计有 243 个, 等等. 因此, 通常只考虑 3^k 设计的单次重复, 并把较高阶的交互作用组合起来作为误差的估计量. 作为说明, 如果三因子交互作用和更高阶的交互作用可被忽略, 则 3^3 设计的单次重复有 8 个误差自由度, 3^4 设计有 48 个误差自由度. 对 $k\geqslant 3$ 个因子说来, 这些设计仍是较大的设计, 因而不太有用.

9.2 3^k 析因设计的混区设计

即使考虑 3^k 设计的单次重复, 由于设计需要做太多的试验, 以至于不大可能使所有的 3^k 次试验都在一致的条件下进行. 这样一来, 经常需要进行混区设计. 3^k 设计可以被混杂在 3^p 个不完全区组中, 其中 $p<k$. 于是, 这些设计可以混杂在 3 个区组中, 9 个区组中等.

9.2.1 三区组的 3^k 析因设计

假定我们想把 3^k 设计混杂在 3 个不完全区组中. 3 个区组有两个自由度, 这样, 必须有两个自由度与区组相混. 在 3^k 析因系列中, 每一主效应都有两个自由度. 每一个二因子交互作用有 4 个自由度而且可分解为两个二自由度的交互作用分量 (例如, AB 和 AB^2), 每一个三因子交互作用有 8 个自由度而且可分解为 4 个二自由度的交互作用分量 (例如, ABC, ABC^2, AB^2C, AB^2C^2), 等等. 因此, 将交互作用分量与区组相混是方便的.

一般的方法是, 构造一**定义对照**

$$L=\alpha_1x_1+\alpha_2x_2+\cdots+\alpha_kx_k \tag{9.2}$$

其中 α_i 代表在相混的效应中第 i 个因子的指数, x_i 是在指定的处理组合中第 i 个因子的水平. 对于 3^k 系列, $\alpha_i=0,1,2$, 但第一个非零组合的 α_i 是 1; $x_i=0$ (低水平), 1 (中水平), 2 (高水平). 3^k 设计的处理组合根据 $L\ (\mathrm{mod}\ 3)$ 的值分派至各区组. 因为 $L\ (\mathrm{mod}\ 3)$ 只能取 0, 1, 2,

所以, 3 个区组被唯一确定. 满足 $L=0 \pmod 3$ 的处理组合组成**主区组** (principal block). 这个区组总是包含处理组合 $00\cdots 0$.

例如, 假定我们想构造一个三区组的 3^2 析因设计. AB 交互作用的两个分量 AB 或 AB^2 都可与区组相混. 比如就取 AB^2, 得定义对照

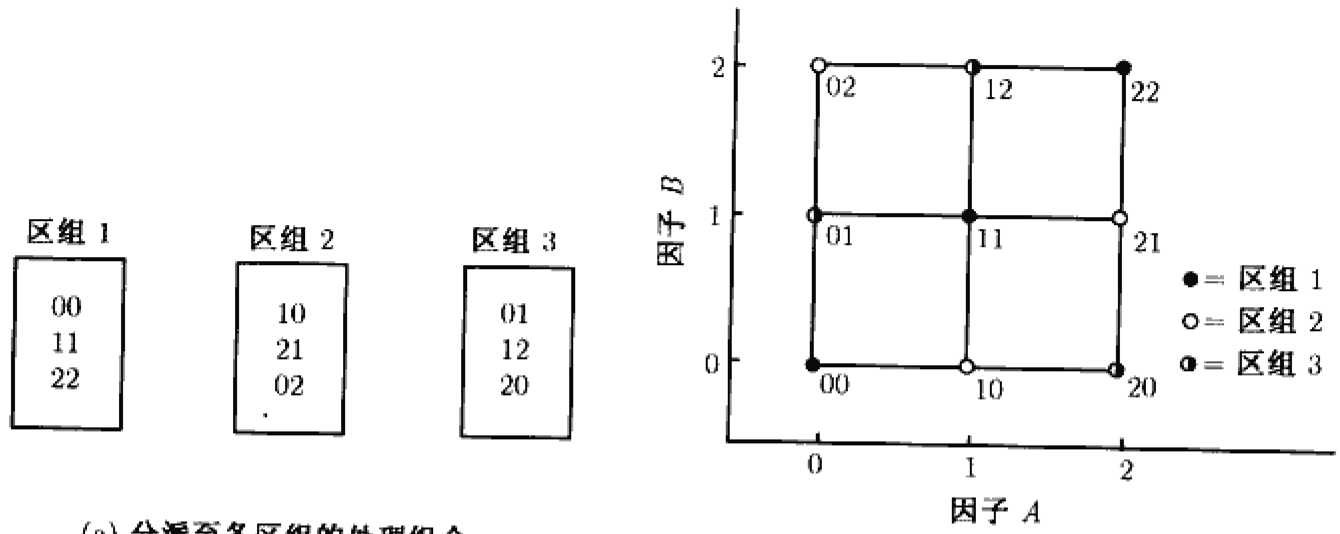
$$L = x_1 + 2x_2$$

求得每一处理组合的 $L \pmod 3$ 的值如下:

$$\begin{aligned} &00: \ L = 1(0)+2(0) = 0 = 0 \pmod 3 \qquad &11: \ L = 1(1)+2(1) = 3 = 0 \pmod 3 \\ &01: \ L = 1(0)+2(1) = 2 = 2 \pmod 3 \qquad &21: \ L = 1(2)+2(1) = 4 = 1 \pmod 3 \\ &02: \ L = 1(0)+2(2) = 4 = 1 \pmod 3 \qquad &12: \ L = 1(1)+2(2) = 5 = 2 \pmod 3 \\ &10: \ L = 1(1)+2(0) = 1 = 1 \pmod 3 \qquad &22: \ L = 1(2)+2(2) = 6 = 0 \pmod 3 \end{aligned}$$

$$20: \ L = 1(2)+2(0) = 2 = 2 \pmod 3$$

各区组见图 9.6.



(a) 分派至各区组的处理组合　　(b) 几何表达

图 9.6　与 AB^2 混杂的三区组 3^2 设计

主区组中的元素形成加法模 3 的一个群. 参见图 9.6, 可得 11+11=22, 11+22=00. 另两个区组的处理组合, 利用加法模 3, 用新区组的任一元素和主区组的元素相加来生成. 这样, 对于区组 2, 使用 10, 得

$$10+00=10 \quad 10+11=21 \quad 10+22=02$$

要生成区组 3, 用 01, 得

$$01+00=01 \quad 01+11=12 \quad 01+22=20$$

例 9.2 使用图 9.6 所示的 3^2 设计的单次重复及所得的下述数据, 说明将 3^2 设计分为 3 个区组的混区设计的统计分析法.

	区组 1	区组 2	区组 3
	00 = 4	10 = −2	01 = 5
	11 = −4	21 = 1	12 = −5
	22 = 0	02 = 8	20 = 0
区组总和 =	0	7	0

用析因设计的常规分析方法, 求得 $SS_A = 131.56$ 和 $SS_B = 0.22$.

还求得

$$SS_{区组} = \frac{(0)^2 + (7)^2 + (0)^2}{3} - \frac{(7)^2}{9} = 10.89$$

然而, $SS_{区组}$ 实际上等于交互作用的 AB^2 分量. 为解释这一点, 将观测值写为:

		因子 B		
		0	1	2
因子 A	0	4	5	8
	1	−2	−4	−5
	2	0	1	0

根据 9.1.2 节, 将上面的数表按左至右的对角线的数计算其和的平方和, 即可求得 AB 交互作用的 I 分量或 AB^2 分量. 这就得出

$$SS_{AB^2} = \frac{(0)^2 + (0)^2 + (7)^2}{3} - \frac{(7)^2}{9} = 10.89$$

它与 $SS_{区组}$ 相同.

方差分析见表 9.4. 因为仅有一次重复, 所以不能进行正规的检验. 注意, 用交互作用的 AB 分量作为误差的估计量并不是好主意.

表 9.4 例 9.2 中数据的方差分析

方差来源	平方和	自由度
区组 (AB^2)	10.89	2
A	131.56	2
B	0.22	2
AB	2.89	2
总和	145.56	8

现在, 看一个稍微复杂一点的设计 ——3^3 析因设计分为 3 个区组, 每个区组有 9 个试验的混区设计. 三因子交互作用的分量 AB^2C^2 用来与区组相混. 该定义的对照是

$$L = x_1 + 2x_2 + 2s_3$$

容易看到, 处理组合 000, 012, 101 属于主区组. 主区组其余的试验如下生成:

(1) 000	(4) $101 + 101 = 202$	(7) $101 + 021 = 122$
(2) 012	(5) $012 + 012 = 021$	(8) $012 + 202 = 211$
(3) 101	(6) $101 + 012 = 110$	(9) $021 + 202 = 220$

为生成另一区组的试验, 注意到 200 不在主区组内. 于是, 区组 2 的元素是

(1) $200 + 000 = 200$	(4) $200 + 202 = 102$	(7) $200 + 122 = 022$
(2) $200 + 012 = 212$	(5) $200 + 021 = 221$	(8) $200 + 211 = 111$
(3) $200 + 101 = 001$	(6) $200 + 110 = 010$	(9) $200 + 220 = 120$

所有这些试验都满足 $L = 2 \pmod 3$. 通过观察发现 100 不属于区组 1 或 2, 可找出最后一个

区组. 像上面一样, 使用 100 得

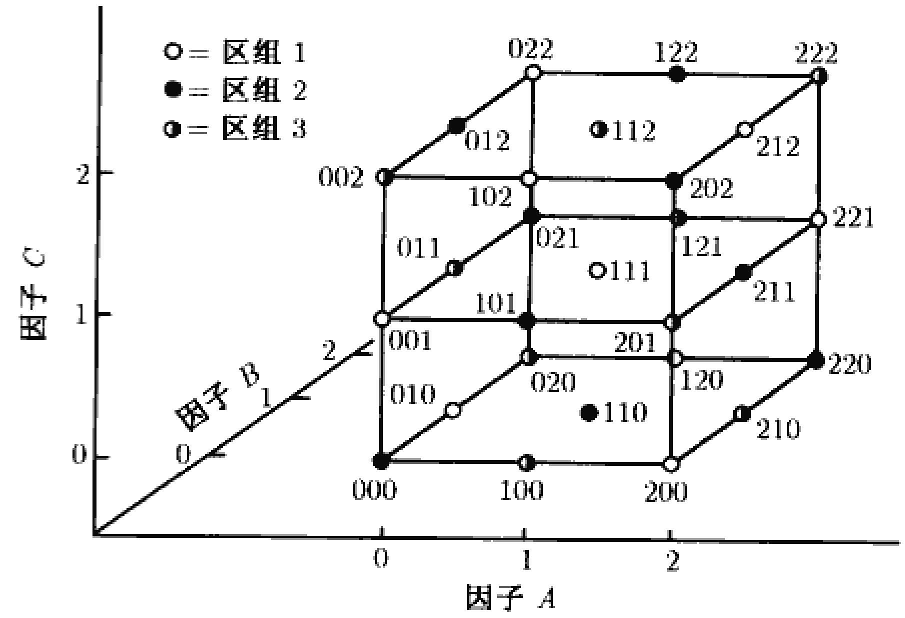
(1) $100+000=100$	(4) $100+202=002$	(7) $100+122=222$
(2) $100+012=112$	(5) $100+021=121$	(8) $100+211=011$
(3) $100+101=201$	(6) $100+110=210$	(9) $100+220=020$

这些区组如图 9.7 所示.

此设计的方差分析见表 9.5. 通过使用这一混区设计方案, 可得到关于所有主效应和二因子交互作用的信息. 三因子交互作用的其余分量 (ABC、AB^2C 和 ABC^2) 组合成误差的估计量. 这 3 个分量的平方和可以用减法求得. 一般来说, 对于三区组的 3^k 混区设计, 总可选择最高阶交互作用的一个分量与区组相混. 这一交互作用的其余的未相混的分量的平方和, 可用普通方法计算出 k 个因子的交互作用的平方和, 再从中减去区组的平方和, 即可求得.

区组 1	区组 2	区组 3
000	200	100
012	212	112
101	001	201
202	102	002
021	221	121
110	010	210
122	022	222
211	111	011
220	120	020

(a) 分派至各区组的处理组合



(b) 几何表达

图 9.7 与 AB^2C^2 混杂的三区组 3^3 设计

表 9.5 与 AB^2C^2 混杂的 3^3 设计的方差分析

方差来源	自由度
区组 (AB^2C^2)	2
A	2
B	2
C	2
AB	4
AC	4
BC	4
误差 ($ABC+AB^2C+ABC^2$)	6
总和	26

9.2.2 九区组的 3^k 析因设计

在有些实验中, 必须进行 9 个区组的 3^k 设计的混区设计. 这样一来, 就有 8 个自由度与区组相混. 要构造出这些设计, 先选取交互作用的两个分量, 从而自动地再得出两个, 得出所需的 8 个自由度. 这自动产生的两个是原来所选的两个效应的广义交互作用. 对 3^k 系统来说, 两个效应 (例如, P 和 Q) 的**广义交互作用**定义为 PQ 和 PQ^2 (或 P^2Q).

原先选取的交互作用的两个分量产生两个定义对照

$$\begin{aligned} L_1 &= \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k = u \pmod 3 \qquad u = 0,1,2 \\ L_2 &= \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k = h \pmod 3 \qquad h = 0,1,2 \end{aligned} \tag{9.3}$$

其中 $\{\alpha_i\}$ 和 $\{\beta_j\}$ 分别是这两个分量的指数, 为方便起见, 第一个非零的 α_i 和 β_j 是 1. (9.3) 式的定义对照隐含了 L_1 和 L_2 的值所指定的 9 个联立方程. 具有 (L_1, L_2) 的同一对值的处理组合被分派在同一区组中.

主区组由满足 $L_1 = L_2 = 0 \pmod 3$ 的处理组合组成. 这个区组的元素关于加法模 3 构成一群. 这样一来, 9.2.1 节给出的方案就可用来生成这些区组.

作为例子, 考虑 3^4 析因设计分为 9 个区组、每区组进行 9 次试验的混区设计. 假定选取 ABC 和 AB^2D^2 混区, 它们的广义交互作用

$$(ABC)(AB^2D^2) = A^2B^3CD^2 = (A^2B^3CD^2)^2 = AC^2D$$
$$(ABC)(AB^2D^2)^2 = A^3B^5CD^4 = B^2CD = (B^2CD)^2 = BC^2D^2$$

亦与区组相混. ABC 和 AB^2D^2 的定义对照是

$$L_1 = x_1 + x_2 + x_3, \qquad L_2 = x_1 + 2x_2 + 2x_4 \tag{9.4}$$

用定义对照 [(9.4) 式] 和主区组的群论性质就可构造出 9 个区组. 此设计见图 9.8.

	区组 1	区组 2	区组 3	区组 4	区组 5	区组 6	区组 7	区组 8	区组 9
	0000	0001	2000	0200	0020	0010	1000	0100	0002
	0122	0120	2122	0022	0112	0102	1122	0222	0121
	0211	0212	2211	0111	0201	0221	1211	0011	0210
	1021	1022	0021	1221	1011	1001	2021	1121	1020
	1110	1111	0110	1010	1100	1120	2110	1210	1112
	1202	1200	0202	1102	1222	1212	2202	1002	1201
	2012	2010	1012	2212	2002	2022	0012	2112	2011
	2101	2102	1101	2001	2121	2111	0101	2201	2100
	2220	2221	1220	2120	2210	2200	0220	2020	2222
$(L_1,L_2)=$	(0,0)	(0,2)	(2,2)	(2,1)	(2,0)	(1,6)	(1,1)	(1,2)	(0,1)

图 9.8 ABC, AB^2D^2, AC^2D, BC^2D^2 与混杂的九区组 3^4 设计

对于九区组的 3^k 混区设计, 有 4 个交互作用分量要被混区. 其余未混区的交互作用分量可以由完整交互作用的平方和减去混区分量的平方和来确定. 9.1.3 节所述的方法可以用来计算交互作用的分量.

9.2.3 3^p 个区组的 3^k 析因设计

3^k 析因设计可以按 3^p 个区组、每个区组有 3^{k-p} 个观测值进行混区设计, 其中 $p < k$. 方法是, 选取 p 个独立效应与区组混区. 从而, $(3^p - 2p - 1)/2$ 个其他的效应即被自动混区. 这些效应是原来选取的那些效应的广义交互作用.

作为说明, 考虑 3^7 设计按 27 个区组的混区设计. 因为 $p = 3$, 我们选取 3 个独立的交互作用分量和 $[3^3 - 2(3) - 1]/2 = 10$ 个其他效应自动混区. 假定选取了 ABC^2DG、BCE^2F^2G 和

$BDEFG$. 由这些效应构造 3 个定义对照并用前面所讲的方法生成 27 个区组. 其他 10 个与区组混区的效应是

$$
\begin{aligned}
&(ABC^2DG)(BCE^2F^2G) = AB^2DE^2F^2G^2\\
&(ABC^2DG)(BCE^2F^2G)^2 = AB^3C^4DE^4F^4G^3 = ACDEF\\
&(ABC^2DG)(BDEFG) = AB^2C^2D^2EFG^2\\
&(ABC^2DG)(BDEFG)^2 = AB^3C^2D^3E^2F^2G^3 = AC^2E^2F^2\\
&(BCE^2F^2G)(BDEFG) = B^2CDE^3F^3G^2 = BC^2D^2G\\
&(BCE^2F^2G)(BDEFG)^2 = B^3CD^2E^4F^4G^3 = CD^2EF\\
&(ABC^2DG)(BCE^2F^2G)(BDEFG) = AB^3C^3D^2E^3F^3G^3 = AD^2\\
&(ABC^2DG)^2(BCE^2F^2G)(BDEFG) = A^2B^4C^5D^3G^4 = AB^2CG^2\\
&(ABC^2DG)(BCE^2F^2G)^2(BDEFG) = ABCD^2E^2F^2G\\
&(ABC^2DG)(BCE^2F^2G)(BDEFG)^2 = ABC^3D^3E^4F^4G^4 = ABEFG
\end{aligned}
$$

这是一个庞大的设计, 需要 $3^7 = 2\ 187$ 个观测值, 分派在 27 个区组中, 每个区组中有 81 个观测值.

9.3 3^k 析因设计的分式重复

分式重复的概念可推广至 3^k 析因设计中. 因为 3^k 设计的完全重复即使对中等的 k 值也需要做很大数目的试验, 所以, 我们感兴趣于这些设计的分式重复. 然而, 正如我们将会看到的, 某些这类设计具有讨厌的别名结构.

9.3.1 3^k 析因设计的 1/3 分式设计

3^k 设计的最大的分式设计是含有 3^{k-1} 个试验的 1/3 分式设计. 因此, 我们把它称为 3^{k-1} 分式析因设计. 要构造一个 3^{k-1} 分式析因设计, 选择一个 2 自由度的交互作用 (一般说来, 是最高阶的交互作用) 的分量, 并把完全 3^k 设计分解为 3 个区组. 3 个区组的每一个都是一个 3^{k-1} 分式设计, 而且任一区组均可被选用. 如果 $AB^{\alpha_2}C^{\alpha_3}\cdots K^{\alpha_k}$ 是用来确定区组的交互作用分量, 则 $I = AB^{\alpha_2}C^{\alpha_3}\cdots K^{\alpha_k}$ 叫做分式析因设计的**定义关系**. 由 3^{k-1} 设计估计的每个主效应或交互效应分量有两个别名, 它们可以用此效应乘以 I 和 I^2 指数模 3 来求得.

作为一个例子, 考虑 3^3 设计的 1/3 分式设计. 选取 ABC 交互作用的任一分量来构造这一设计, 也就是选 ABC、AB^2C、ABC^2 或 AB^2C^2. 这样, 3^3 设计实际上有 12 种不同的 1/3 分式设计, 它们由

$$x_1 + \alpha_2x_2 + \alpha_3x_3 = u \pmod 3$$

确定, 其中 $\alpha_i = 1$ 或 $2, u = 0$、1 或 2. 假定我们选取分量 AB^2C^2. 所得的每个 3^{3-1} 分式设计将刚好含有 $3^2 = 9$ 个处理组合, 它们必须满足

$$x_1 + 2x_2 + 2x_3 = u \pmod 3$$

其中 $u = 0$、1 或 2. 容易证实, 3 个 1/3 分式设计如图 9.9 所示.

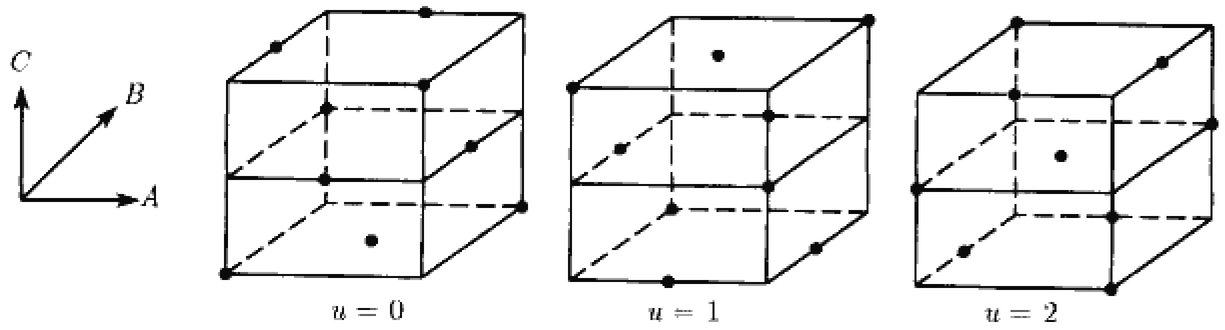
如果进行了图 9.9 中任一种 3^{3-1} 设计的实验, 则所得的别名结构是

$$
\begin{aligned}
&A = A(AB^2C^2) = A^2B^2C^2 = ABC, && A = A(AB^2C^2)^2 = A^3B^4C^4 = BC \\
&B = B(AB^2C^2) = AB^3C^2 = AC^2, && B = B(AB^2C^2)^2 = A^2B^5C^4 = ABC^2 \\
&C = C(AB^2C^2) = AB^2C^3 = AB^2, && C = C(AB^2C^2)^2 = A^2B^4C^5 = AB^2C \\
&AB = AB(AB^2C^2) = A^2B^3C^2 = AC, && AB = AB(AB^2C^2)^2 = A^3B^5C^4 = BC^2
\end{aligned}
$$

因此, 设计的 8 个自由度实际估计的 4 个效应是 $A+BC+ABC$、$B+AC^2+ABC^2$、$C+AB^2+AB^2C$ 以及 $AB+AC+BC^2$. 此设计仅当所有的交互作用小于主效应时才有实际价值. 由于主效应的别名是二因子交互作用, 这是一个分辨度为III的设计. 注意, 这一设计的别名关系很复杂. 每个主效应的别名都是交互作用的分量. 例如, 如果二因子交互作用 BC 较大, 就可能使 A 的主效应的估计量失真, 并使得 $AB+AC+BC^2$ 效应极难解释. 除非假定所有交互作用都可忽略, 否则, 我们很难看出怎样应用这一设计.

设计 1 $u=0$	设计 2 $u=1$	设计 3 $u=2$
000	100	200
012	112	212
101	201	011
202	002	102
021	121	221
110	210	010
122	222	022
211	011	111
220	020	120

(a) 处理组合



(b) 几何意义

图 9.9　定义关系为 $I = AB^2C^2$ 的 3^3 设计的 3 个 1/3 分式设计

在离开 3_{III}^{3-1} 设计之前, 注意到对于设计 $u=0$ (见图 9.9), 如果令 A 表示行, B 表示列, 则此设计可记为

000	012	021
101	110	122
202	211	220

这是一个 3×3 拉丁方. 对 3_{III}^{3-1} 设计作出唯一解释所需作的忽略交互作用的假定与拉丁方设计的假设是类似的. 不过, 这两种设计的目的不同, 一种是作为分式重复实验的推论, 另一种是来自随机化约束. 由表 4.13 看到, 恰有 12 个 3×3 拉丁方, 而且每一个拉丁方对应于 12 种不同的 3^{3-1} 分式析因设计之一.

定义关系为 $I = AB^{\alpha_2}C^{\alpha_3}\cdots K^{\alpha_k}$ 的 3^{k-1} 设计的处理组合, 可以利用和 2^{k-p} 系统所用的类似方法来构造. 首先用通常的 0, 1, 2 记号写出 $k-1$ 个因子的**完全**三水平析因设计的 3^{k-1} 个试验. 就是第 8 章所说的**基本设计**. 然后写出第 k 个因子, 令它的水平 x_k 等于最高阶交互作用的合适的分量, 比方说 $AB^{\alpha_2}C^{\alpha_3}\cdots(K-1)^{\alpha_{k-1}}$, 所给出的关系式

$$x_k = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-1}x_{k-1} \tag{9.5}$$

这里 $\beta_i = (3-\alpha_k)\alpha_i \pmod 3$, $1 \leqslant i \leqslant k-1$. 这就产生一个有最大可能的分辨度的设计.

作为说明, 我们用这一方法生成定义关系为 $I = AB^2CD$ 的 3_{IV}^{4-1} 设计, 如表 9.6 所示. 容易证实, 此表中每一处理组合的前 3 个数字就是完全 3^3 设计的 27 个试验. 这就是基本设计. 对 AB^2CD, 我们有 $\alpha_1=\alpha_3=\alpha_4=1$, $\alpha_2=2$. 所以, $\beta_1=(3-1)\alpha_1(\text{mod } 3)=(3-1)(1)=2$, $\beta_2=(3-1)\alpha_2(\text{mod } 3)=(3-1)(2)=4=1(\text{mod } 3)$, $\beta_3=(3-1)\alpha_3(\text{mod } 3)=(3-1)(1)=2$, 于是, (9.5) 式变成

$$x_4 = 2x_1 + x_2 + 2x_3 \tag{9.6}$$

第 4 个因子的水平满足 (9.6) 式. 例如, $2(0)+1(0)+2(0)=0$, $2(0)+1(1)+2(0)=1$, $2(1)+1(1)+2(0)=3=0$, 等等.

表 9.6 $I = AB^2CD$ 的 3_{IV}^{4-1} 设计

0000	0012	2221
0101	0110	0021
1100	0211	0122
1002	1011	0220
0202	1112	1020
1201	1210	1121
2001	2010	1222
2102	2111	2022
2200	2212	2120

所得的 3_{IV}^{4-1} 设计有 26 个自由度, 可以用于计算 13 个主效应和交互作用分量 (及其别名) 的平方和. 任一效应的别名可以用通常的方法求得, 例如, A 的别名是 $A(AB^2CD)=ABC^2D^2$ 和 $A(AB^2CD)^2=BC^2D^2$. 可以证实, 4 个主效应和任一二因子交互作用分量分离开来, 但某些二因子交互作用分量彼此互为别名. 我们再次看到别名结构的复杂性. 即使有一个二因子交互作用较大, 也可能很难将它们从设计中分离出来.

3^{k-1} 设计的统计分析用通常的析因实验方差分析法来完成. 交互作用分量平方和的计算和 9.1 节相同. 在解释结果时, 交互作用的分量没有实际意义.

9.3.2 其他的 3^{k-p} 分式析因设计

对中等大的 k 值, 可以考虑 3^k 设计更小分式的分式设计. 一般说来, 可以构造 3^k 设计的 $(\frac{1}{3})^p$ 分式设计, $p<k$, 此分式设计含有 3^{k-p} 个试验. 这样的设计叫做 3^{k-p} 分式析因设计. 于是, 3^{k-2} 设计就是 1/9 分式设计, 3^{k-3} 设计是 1/27 分式设计, 等等.

构造 3^{k-p} 分式析因设计的方法是, 选取交互作用的 p 个分量, 用这些效应将 3^k 个处理组合分解为 3^p 个区组. 于是, 每一个区组是一个 3^{k-p} 分式析因设计. 任一分式设计的定义关系 I 是由原来选取的 p 个效应和它们的 $(3^p-2p-1)/2$ 个广义交互作用所组成. 任一主效应或交互作用分量的别名由效应乘以 I 和 I^2 指数模为 3 而求得.

也可以按 9.3.1 节那样做, 先写出完全 3^{k-p} 析因设计的处理组合, 然后引进附加的 p 个因子, 令它们等于交互作用的分量, 来生成确定的 3^{k-p} 分式析因设计的试验.

我们通过构造一个 3^{4-2} 设计, 即 3^4 设计的 1/9 分式设计, 来说明这一方法. 令 AB^2C 和 BCD 是用来构造这一设计所选取的两个交互作用分量. 它们的广义交互作用是 $(AB^2C)(BCD) = AC^2D$ 和 $(AB^2C)(BCD)^2 = ABD^2$. 于是, 此设计的定义关系是 $I = AB^2C = BCD = AC^2D = ABD^2$, 而且是分辨度为III的. 写出因子 A 和 B 的 3^2 设计, 然后加上由

$$x_3 = 2x_1 + x_2, \qquad x_4 = 2x_2 + 2x_3$$

得出的两个新的因子, 就求得此设计的 9 个处理组合. 这等价于用 AB^2C 和 BCD 把完全的 3^4 设计分解为 9 个区组, 然后选取这些区组之一作为所要求的分式设计. 完整的设计如表 9.7 所示.

这一设计有 8 个自由度, 可用于估计 4 个主效应和它们的别名. 任一效应的别名的求法, 是将此效应乘以 AB^2C, BCD, AC^2D, ABD^2 及它们的平方, 取指数模 3 即可. 此设计的完整别名结构见表 9.8.

表 9.7 $I = AB^2C$ 和 $I = BCD$ 的 3^{4-2}_{III} 设计

0000	0111	0222
1021	1102	1210
2012	2120	2201

表 9.8 表 9.7 中的 3^{4-2}_{III} 设计的别名结构

效应	别名							
	I				I^2			
A	ABC^2	$ABCD$	ACD^2	AB^2D	BC^2	$AB^2C^2D^2$	CD^2	BD^2
B	AC	BC^2D^2	ABC^2D	AB^2D^2	ABC	CD	AB^2C^2D	AD^2
C	AB^2C^2	BC^2D	AD	$ABCD^2$	AB^2	BD	ACD	ABC^2D^2
D	AB^2CD	BCD^2	AC^2D^2	AB	AB^2CD^2	BC	AC^2	ABD

由别名结构可见, 这一设计仅当没有交互作用时才有用. 还有, 如果将 A 记为行, B 记为列, 则检查一下表 9.7 即可知, 该 3^{4-2}_{III} 设计也是一个正交拉丁方.

对于 $4 \leqslant k \leqslant 10$, Connor and Zelen (1959) 中有众多可供选择的设计, 这本小册子是提供给国家标准局的, 有 3^{k-p} 设计最完全的表.

在本节中, 我们多次指出过 3^{k-p} 分式析因设计别名关系的复杂性. 一般说来, 当 k 中等大时, 比方说 $k \geqslant 4$ 或 5 时, 3^k 设计的大小迫使大多数实验者去考虑较小的分式设计. 不幸的是, 这些设计的别名关系涉及**部分**二自由度的交互作用分量作为别名. 设计中的这一结果, 反过来又给分析带来困难, 如果交互作用不可忽略的话, 就难于对数据作出解释. 还有, 没有简单的扩编方案 (如同折叠那样) 能够用来把两个或多个分式设计组合起来, 以便把显著的交互作用分离开来. 当出现弯曲时, 常认为 3^k 设计是合适的, 不过, 可能有更有效的其他方法 (参阅第 11 章). 正因为如此, 我们认为 3^{k-p} 分式析因设计作为解决问题的解法说来, 通常不是好的设计.

9.4　混合水平的析因设计

我们着重研究了所有因子水平数相同的析因设计和分式析因设计. 第 6~8 章讨论的二水平系统特别有用. 本章前面提到的三水平系统不太用到, 因为即使是对一般的因子个数来说, 此种设计也相对较大, 而且大多数较小的分式设计的别名关系又很复杂, 需要对交互作用作出严格的假定才可使用.

我们相信, 二水平的析因设计和分式设计将成为工业实验的柱石, 用于产品和工序的开发、改进和故障探测等. 但是, 有时需要包含多于两个水平的一个因子 (或几个因子). 这通常发生在实验中同时有定量因子和定性因子, 且定性因子有 3 个水平的情形. 若所有因子都是定量因子, 应选用具有中心点的二水平设计. 本节将指出, 怎样把一些三水平和四水平的因子收容在 2^k 设计中.

9.4.1　二水平和三水平的因子

我们偶尔想要研究有些因子是二水平、一些因子是三水平的设计. 如果这个设计是完全析因设计, 那么此设计的构造与分析没有任何新的方法. 然而, 若我们正在考虑一个分式析因设计, 那么就会对这些设计感兴趣. 如果所有因子是定量因子, 则用混合水平的分式设计代替带有中心点的 2^{k-p} 分式析因设计通常并不好. 通常在考虑这些设计时, 实验者需同时考虑定性因子和定量因子, 其中定性因子是三水平的. 在 3^{k-p} 设计中, 我们看到的定性因子的复杂别名问题大都带进了混合水平分式系统中. 全部为定性因子或定性因子与定量因子混合的混合水平分式设计都应谨慎使用. 本节主要讨论这些设计中的一部分.

一些因子有 2 个水平而另一些因子有 3 个水平的设计通常可以用 2^k 设计的加减符号表推导出来. 一般方法最好用例子来说明. 假定有两个变量, A 有 2 个水平, X 有 3 个水平. 考虑有 8 个试验的 2^3 设计加减符号表. 表 9.9 左边列出了列 B 和 C 的符号. 令 X 的水平表示为 x_1, x_2, x_3. 表 9.9 的右边表示如何将 B 和 C 的符号组合起来形成三水平因子的水平.

表 9.9　用二水平因子来构造三水平因子

二水平因子		三水平因子
B	C	X
−	−	x_1
+	−	x_2
−	+	x_2
+	+	x_3

现在, 因子 X 有两个自由度, 如果该因子是定量的, 则可将其分解为一个线性分量和一个二次分量, 每个分量有一个自由度. 表 9.10 表示了这一 2^3 设计, 列的名称表示需估计的实际效应, X_L 和 X_Q 分别表示 X 的线性效应和二次效应. X 的线性效应是两个效应估计量的和, 由与 B 和 C 有关的列算得, A 的效应只可以由 X 处于低水平或高水平的试验 (即试验 1, 2, 7, 8) 来计算. 同理, $A\times X_L$ 效应是两个效应的和, 可由 AB 列和 AC 列算得. 又, 试验 3 和试验 5 是重复试验. 因此, 单自由度的误差估计量可以用这两个试验求得. 同理, 试验 4 和试验 6 是重复试验, 它可导出第 2 个单自由度的误差估计量. 这两对试验的平均方差可用作二自由度的

误差均方. 完整的方差分析见表 9.11.

表 9.10 一个二水平因子和一个三水平因子的 2^3 设计

	A	X_L	X_L	$A\times X_L$	$A\times X_L$	X_Q	$A\times X_Q$	实际的处理组合	
试验	A	B	C	AB	AC	BC	ABC	A	X
1	−	−	−	+	+	+	−	低	低
2	+	−	−	−	−	+	+	高	低
3	−	+	−	−	+	−	+	低	中
4	+	+	−	+	−	−	−	高	中
5	−	−	+	+	−	−	+	低	中
6	+	−	+	−	+	−	−	高	中
7	−	+	+	−	−	+	−	低	高
8	+	+	+	+	+	+	+	高	高

如果能够假定二因子交互作用和更高阶交互作用可被忽略, 则可以把表 9.10 的设计转换为至多有 4 个二水平因子和 1 个三水平因子的分辨度为III的分式设计. 这一点可以把二因子水平的列 A、AB、AC 和 ABC 结合起来完成. 列 BC 不能用于二水平因子, 因为它含有三水平因子 X 的二次效应.

表 9.11 表 9.10 中的设计的方差分析

方差来源	平方和	自由度	均方
A	SS_A	1	MS_A
$X(X_L+X_Q)$	SS_X	2	MS_X
$AX(A\times X_L+A\times X_Q)$	SS_{AX}	2	MS_{AX}
误差 (来自试验 3 与试验 5 以及试验 4 与试验 6)	SS_E	2	MS_E
总和	SS_T	7	

同样的方法可应用到有 16 个试验、32 个试验和 64 个试验的 2^k 设计中. 对 16 个试验说来, 可以构造有 2 个二水平因子和 2 个或 3 个三水平因子的分辨度为V的分式析因设计. 也可以得到有 3 个二水平因子和 1 个三水平因子的 16 个试验的分辨度为V的分式设计. 如果 16 个试验包含 4 个二水平因子和 1 个三水平因子, 则设计的分辨度为III. 可用同样的方法设计 32 个试验和 64 个试验的设计. 这些设计的一些其他方面的讨论, 参阅 Addleman (1962).

9.4.2 二水平和四水平的因子

非常容易将四水平因子安置在 2^k 设计中. 方法就是用 2 个二水平因子代表四水平因子. 例如, 设 A 是一个四水平因子, 其水平为 a_1, a_2, a_3, a_4. 考虑通常的加减符号表中的两列, 比方说列 P 和 Q. 这两列的符号如表 9.12 的左边所示. 此表的右边表示怎样将这些符号与因子 A 的 4 个水平相对应. 由列 P 和 Q 与 PQ 交互作用代表的效应是相互正交的, 且对应于 3 个自由度的 A 效应. 由 2 个二水平因子构造 1 个四水平因子的这种方法称为**替换法** (replacement).

为了更全面地说明这一思想, 假定有 1 个四水平的因子和 2 个二水平的因子, 并且要估计所有的主效应和涉及这些因子的交互作用. 使用 16 个试验的设计就能够做到这一点. 表 9.13 是通常有 16 个试验的 2^4 设计加减符号表, 列 A 和列 B 用来形成四水平因子, 比方说 X, 其水平为 x_1, x_2, x_3, x_4. 与通常的 2^k 系统一样, 计算每一列 $A, B, \cdots, ABCD$ 的平方和. 因子 X, C, D 以及它们的交互作用的平方和如下:

表 9.12 四水平因子 A 表示为 2 个二水平因子

试验	二水平因子		四水平因子
	P	Q	A
1	−	−	a_1
2	+	−	a_2
3	−	+	a_3
4	+	+	a_4

$$SS_X = SS_A + SS_B + SS_{AB} \quad (3\text{ 个自由度})$$
$$SS_C = SS_C \quad (1\text{ 个自由度})$$
$$SS_D = SS_D \quad (1\text{ 个自由度})$$
$$SS_{CD} = SS_{CD} \quad (1\text{ 个自由度})$$
$$SS_{XC} = SS_{AC} + SS_{BC} + SS_{ABC} \quad (3\text{ 个自由度})$$
$$SS_{XD} = SS_{AD} + SS_{BD} + SS_{ABD} \quad (3\text{ 个自由度})$$
$$SS_{XCD} = SS_{ACD} + SS_{BCD} + SS_{ABCD} \quad (3\text{ 个自由度})$$

它可称为 4×2^2 设计. 如果想要略去二因子的交互作用, 则在二因子交互作用 (除了 AB 之外) 列. 三因子的交互作用列和四因子交互作用列共可安排至多 9 个附加的二水平因子.

表 9.13 1 个四水平因子和 2 个二水平因子的 16 个试验

试验	(A	B)	$=X$	C	D	AB	AC	BC	ABC	AD	BD	ABD	CD	ACD	BCD	$ABCD$
1	−	−	x_1	−	−	+	+	+	−	+	+	−	+	−	−	+
2	+	−	x_2	−	−	−	−	+	+	−	+	+	+	+	−	−
3	−	+	x_3	−	−	−	+	−	+	+	−	+	+	−	+	−
4	+	+	x_4	−	−	+	−	−	−	−	−	−	+	+	+	+
5	−	−	x_1	+	−	+	−	−	+	+	+	−	−	+	+	−
6	+	−	x_2	+	−	−	+	−	−	−	+	+	−	−	+	+
7	−	+	x_3	+	−	−	−	+	−	+	−	+	−	+	−	+
8	+	+	x_4	+	−	+	+	+	+	−	−	−	−	−	−	−
9	−	−	x_1	−	+	+	+	+	−	−	−	+	−	+	+	−
10	+	−	x_2	−	+	−	−	+	+	+	−	−	−	−	+	+
11	−	+	x_3	−	+	−	+	−	+	−	+	−	−	+	−	+
12	+	+	x_4	−	+	+	−	−	−	+	+	+	−	−	−	−
13	−	−	x_1	+	+	+	−	−	+	−	−	+	+	−	−	+
14	+	−	x_2	+	+	−	+	−	−	+	−	−	+	+	−	−
15	−	+	x_3	+	+	−	−	+	−	−	+	−	+	−	+	−
16	+	+	x_4	+	+	+	+	+	+	+	+	+	+	+	+	+

二水平因子与四水平因子混合的分式析因设计有广泛的适用范围. 然而, 我们劝告大家要慎重地使用这些设计. 如果所有的因子是定量的, 带有中心点的 2^{k-p} 系统是更好的选择. 当实验中同时有定量因子与定性因子时, 如果要求有二水平因子与四水平因子的设计的分辨度达到IV或更高, 则设计通常会相当大, 多数情况下需要 $n \geqslant 32$ 次试验.

9.5 思 考 题

*9.1 研究显影液浓度 (A) 和冲洗时间 (B) 对底片密度的效应. 用 3 种浓度和 3 种时间, 进行有 4 次重复的 3^2 析因设计. 所得的数据如下. 用析因实验的标准方法分析这些数据.

显影液浓度	冲洗时间 (分)					
	10		14		18	
1	0	2	1	3	2	5
	5	4	4	2	4	6
2	4	6	6	8	9	10
	7	5	7	7	8	5
3	7	10	10	10	12	10
	8	7	8	7	9	8

9.2 计算思考题 9.1 的二因子交互作用的 I 分量和 J 分量.

*9.3 进行一项实验, 涉及 3 种不同类型的 32 盎司瓶子 (A) 和 3 种不同类型的架子 (B)—— 光滑固定架、带网格的狭长陈列架以及饮料冷却器, 要研究把 10 个 12 瓶装的箱子放在架子上所需的时间. 3 名工人 (因子 C) 在实验中工作, 用有 2 次重复的 3^3 析因设计. 观测到的时间数据如下表所示. 分析这些数据并写出结论.

工人	瓶子类型	重复 I			重复 II		
		固定架	网格阵	冷却器	固定架	网格阵	冷却器
	塑料	3.45	4.14	5.80	3.36	4.19	5.23
1	28 mm 玻璃	4.07	4.38	5.48	3.52	4.26	4.85
	38 mm 玻璃	4.20	4.26	5.67	3.68	4.37	5.58
	塑料	4.80	5.22	6.21	4.40	4.70	5.88
2	28 mm 玻璃	4.52	5.15	6.25	4.44	4.65	6.20
	38 mm 玻璃	4.96	5.17	6.03	4.39	4.75	6.38
	塑料	4.08	3.94	5.14	3.65	4.08	4.49
3	28 mm 玻璃	4.30	4.53	4.99	4.04	4.08	4.59
	38 mm 玻璃	4.17	4.86	4.85	3.88	4.48	4.90

9.4 一位药物研究者研究利多卡因对警犬心肌中酶水平的效应. 使用 3 种不同商标的利多卡因 (A)、3 种剂量水平 (B) 和 3 只警犬 (C) 做实验, 进行有 2 次重复的 3^3 析因设计. 观测到的酶水平如下表. 分析此实验所得的这些数据.

利多卡因商标	剂量浓度	重复 I			重复 II		
		警犬			警犬		
		1	2	3	1	2	3
	1	96	84	85	84	85	86
1	2	94	99	98	95	97	90
	3	101	106	98	105	104	103
	1	85	84	86	80	82	84
2	2	95	98	97	93	99	95
	3	108	114	109	110	102	100
	1	84	83	81	83	80	79
3	2	95	97	93	92	96	93
	3	105	100	106	102	111	108

9.5 计算例 9.1 中二因子交互作用的 I 分量和 J 分量.

9.6 在某一化工工艺中, 用 3^2 析因设计做一个实验. 设计的两个因子分别是温度和压力, 响应变量是产率. 实验数据如下:

温度 (℃)	压力 (psig)		
	100	120	140
80	47.58, 48.77	64.97, 69.22	80.92, 72.60
90	51.86, 82.43	88.47, 84.23	93.95, 88.54
100	71.18, 92.77	96.57, 88.72	76.58, 83.04

(a) 用导出方差分析表的方式分析这个实验的数据. 可以得到什么结论.

(b) 用图形方法分析残差. 模型潜在的假定或模型合适性有什么问题吗?

(c) 设两个因子的高、中、低水平为 −1, 0, +1, 证明用最小二乘法求得的产率的二阶模型是

$$\hat{y} = 86.81 + 10.4x_1 + 8.42x_2 - 7.17x_1^2 - 7.84x_2^2 - 7.69x_1x_2$$

(d) 证明 (c) 中的模型用自然变量温度 (T) 和压力 (P) 表示是

$$\hat{y} = -1\ 335.63 + 18.56T + 8.59P - 0.072T^2 - 0.019\ 6P^2 - 0.038\ 4TP$$

(e) 构造作为压力和温度的函数的产率的等高线图. 根据图形, 你可以推荐该化工工艺的好的生产条件吗?

9.7 (a) 用三因子交互作用的 ABC^2 分量, 分为 3 个区组, 对 3^3 设计进行混区设计. 将你的结果与图 9.7 的设计进行比较.

(b) 用三因子交互作用的 AB^2C 分量, 分为 3 个区组, 对 3^3 设计进行混区设计. 将你的结果与图 9.7 的设计进行比较.

(c) 用三因子交互作用的 ABC 分量, 分为 3 个区组, 对 3^3 设计进行混区设计. 将你的结果与图 9.7 的设计进行比较.

(d) 比较 (a)、(b)、(c) 及图 9.7 的结果, 可以得出什么结论?

9.8 用四因子交互作用的 AB^2CD 分量, 分为 3 个区组, 对 3^4 设计进行混区设计.

9.9 考虑思考题 9.3 第一个重复的数据. 假定不能在同一天内做完所有 27 个观测值, 建立一个以 AB^2C 混区, 分为 3 天进行的实验设计. 分析这些数据.

9.10 写出分为 9 个区组的 3^4 设计的方差分析表. 这是实用的设计吗?

9.11 考虑思考题 9.3 的数据. 如果在重复 I 中 ABC 混区而且在重复 II 中 ABC^2 混区, 请进行方差分析.

9.12 考虑思考题 9.3 重复 I 的数据, 假定只进行此设计的以 $I = ABC$ 的 1/3 分式设计实验. 构造这一设计, 确定别名结构, 并分析这些数据.

9.13 观察图 9.9, 如果完成了前 9 个试验, 并删除 3 个因子中的一个因子, 那么会留下哪种类型的设计?

9.14 构造一个 $I = ABCD$ 的 3_{IV}^{4-1} 设计. 写出此设计的别名结构.

9.15 证明思考题 9.14 的设计是分辨度为IV的设计.

9.16 构造一个 $I = ABC$ 和 $I = CDE$ 的 3^{5-2} 设计. 写出此设计的别名结构. 此设计的分辨度是几?

9.17 构造一个 3^{9-6} 设计, 并证明它是分辨度为III的设计.

9.18 构造一个分为两个区组, 每个区组有 16 个观测值的 4×2^3 设计的混区设计. 对此设计写出方差分析表.

9.19 写出 $2^2 3^2$ 析因设计的方差分析表. 讨论怎样将这一设计进行混区设计.

9.20 从 16 个试验的 2^4 设计出发, 指出在此实验中怎样将 2 个三水平因子安排进去. 如果需要一些二因子交互作用的信息, 那么能够安排几个二水平因子?

9.21 从 16 个试验的 2^4 设计出发, 指出怎样安排 1 个三水平因子和 3 个二水平因子, 使人们仍然可以估计二因子交互作用.

9.22 在思考题 8.27 中, 你遇到作者的两位朋友 Harry 和 Judy, 他们在俄勒冈州的 Newberg 有家葡萄酒厂和一座葡萄园. 那个问题描述了如何对他们的 1985 年黑比诺葡萄酒产品应用二水平分式析因设计. 在 1987 年, 他们需要进行另一种黑比诺葡萄酒实验. 此实验的变量是

变　量	水　平
黑比诺无性系	威登斯维尔, 玻玛
浆果大小	小, 大
发酵温度	80°F, 85°F, 90/80°F, 90°F
整串浆果	无, 10%
浸渍时间	10 天, 21 天
酵母类型	阿斯曼型, 香槟型
橡木类型	特朗赛型, 阿里耶型

Harry 和 Judy 决定用 16 个试验的二水平分式析因设计, 把四水平的发酵温度作为 2 个二水平变量来处理. 与思考题 8.27 一样, 他们用品尝小组评出的名次作为响应变量. 设计以及所得的平均名次如下:

试验	无性系	浆果大小	发酵温度		整串浆果	浸渍时间	酵母类型	橡木类型	平均名次
1	−	−	−	−	−	−	−	−	4
2	+	−	−	−	−	+	+	+	10
3	−	+	−	−	+	−	+	+	6
4	+	+	−	−	+	+	−	−	9
5	−	−	+	−	+	+	+	−	11
6	+	−	+	−	+	−	−	+	1
7	−	+	+	−	−	+	−	+	15
8	+	+	+	−	−	−	+	−	5
9	−	−	−	+	+	+	−	+	12
10	+	−	−	+	+	−	+	−	2
11	−	+	−	+	−	+	+	−	16
12	+	+	−	+	−	−	−	+	3
13	−	−	+	+	−	−	+	+	8
14	+	−	+	+	−	+	−	−	14
15	−	+	+	+	+	−	−	−	7
16	+	+	+	+	+	+	+	+	13

(a) 描述此设计的别名.

(b) 分析这些数据并写出结论.

(c) 关于这个实验和思考题 8.27 的 1985 年黑比诺葡萄酒实验, 你能做什么比较?

9.23 在 W. D. Baten 发表在*Industrial Quality Control* (1956 年卷) 的一篇文章中, 介绍了 3 个因子对钢筋长度的效应的一个实验的研究结果. 每根钢筋要经过两个热处理过程, 并在一天中 3 个时间点 (上午 8 点, 上午 11 点, 下午 3 点) 的某一时间点, 被 4 台机器中的一台切割. 长度的规范数据如下:

(a) 假定每一单元格中的 4 个数据是重复数据, 分析这个实验的数据.

(b) 分析这个实验的残差. 有迹象表明在某个单元中存在异常点吗? 如果找到异常点, 删除异常点后重复 (a) 的分析. 你得到什么结论?

(c) 假设各个单元中的观测值是在 4 台机器上一起进行热处理的钢筋依次被切割的 (规范) 长度. 分

析这些数据并确定 3 个因子对平均长度的效应.

(d) 计算每一单元格中 4 个观测值方差的对数. 分析这个响应. 可以得到什么结论?

(e) 假设钢筋被切割的时间在日常生产过程中是不受控制的. 分析在每个机器/热处理过程中 12 个钢筋长度数据的平均值和方差的对数. 可以得到什么结论?

时间	热处理过程	机器							
		1		2		3		4	
上午 8 点	1	6	9	7	9	1	2	6	6
		1	3	5	5	0	4	7	3
	2	4	6	6	5	−1	0	4	5
		0	1	3	4	0	1	5	4
上午 11 点	1	6	3	8	7	3	2	7	9
		1	−1	4	8	1	0	11	6
	2	3	1	6	4	2	0	9	4
		1	−2	1	3	−1	1	6	3
下午 3 点	1	5	4	10	11	−1	2	10	5
		9	6	6	4	6	1	4	8
	2	6	0	8	7	0	−2	4	3
		3	7	10	0	4	−4	7	0

第 10 章　拟合回归模型

本章纲要

10.1　引言
10.2　线性回归模型
10.3　线性回归模型的参数估计
10.4　多元回归的假设检验
10.4.1　回归的显著性检验
10.4.2　回归系数的个别检验和分组检验
10.5　多元回归的置信区间
10.5.1　单个回归系数的置信区间
10.5.2　平均响应的置信区间
10.6　新响应观测的预测
10.7　回归模型的诊断
10.7.1　尺度残差和 PRESS
10.7.2　影响诊断
10.8　拟合不足检验
第 10 章补充材料
S10.1　回归系数的协方差矩阵
S10.2　回归模型与所设计的实验
S10.3　调整的 R^2
S10.4　逐步回归和其他回归模型变量选择法
S10.5　预测响应的方差
S10.6　预测误差的方差
S10.7　回归模型的杠杆值

10.1　引　　言

很多问题中, 有两个或多个变量是相关的, 我们想要建立模型并探索这些关系. 例如, 在化工工艺中, 产品的产率和运行温度有关. 化学工程师想要建立产率和温度的模型, 并用这一模型进行预测、过程优化和过程控制.

一般地, 设有一个**因变量**或**响应** y, 它依赖于 k 个**自变量**或**回归变量**, 例如 $x_1, x_2, \cdots, x_k$. 这些变量之间的关系可以用一个称为**回归模型**的数学模型来刻画. 它拟合一组样本数据. 在有些场合, 实验者知道 y 和 $x_1, x_2, \cdots, x_k$ 之间真实函数关系的精确表达式, 比方说, $y = \phi\ (x_1, x_2, \cdots, x_k)$. 但是, 大多数情况下, 真实的函数关系是未知的, 实验者要选取一个合适的函数形式来逼近 ϕ . 低阶的多项式模型是广为应用的逼近函数.

在实验设计与回归分析之间有强烈的相互影响. 本书始终强调以**经验模型**定量地表示实验结果的重要性, 这将使模型易于理解、解释和实现. 回归模型是它的基础. 在许多场合, 我们已经给出了描述实验结果的回归模型. 本章将介绍有关拟合这些模型的一些内容. 关于回归的更完整的叙述见 Montgomery, Peck, and Vining (2001) 以及 Myers (1990).

回归分析常用于分析**无计划实验**所得的数据, 例如, 观测不可控现象所得的数据或历史记录. 回归分析在出现某些问题的设计过的实验中也是很有用的. 本章将阐述这些情形.

10.2　线性回归模型

我们将主要拟合线性回归模型. 例如, 假定我们想要建立聚合物的黏度与温度和催化剂进料速率之间的经验模型. 可以描述这个关系的一个模型是

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{10.1}$$

其中 y 表示黏度, x_1 表示温度, x_2 表示催化剂进料速率. 这是有两个自变量的**多元线性回归模型**. 自变量也常称为**预测变量** (predictor varible) 或**回归变量** (regressor). 使用术语**线性**是因为 (10.1) 式是未知参数 β_0, β_1, β_2 的一个线性函数. 模型描述了在二维 x_1, x_2 空间中的一个平面. 参数 β_0 定义为平面的截距. 有时也称 β_1, β_2 为**偏回归系数**, 这是因为 β_1 度量了在 x_1 每改变一个单位而 x_2 保持不变时 y 的期望改变量, 同理, β_2 度量了在 x_2 每改变一个单位而 x_1 保持不变时 y 的期望改变量.

一般地, 响应变量 y 与 k 个回归变量相关. 称模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{10.2}$$

为有 k 个回归变量的**多元线性回归模型**. 称参数 β_j, $j = 0, 1, \cdots, k$, 为**回归系数** (regression coefficient). 这个模型描述了 k 维回归变量$\{x_j\}$ 空间中的超平面. 参数 β_j 表示在 x_j 每改变一个单位而其余自变量 $x_i (i \neq j)$ 保持不变时响应 y 的期望改变量.

比 (10.2) 式看起来更复杂的模型通常也可以用多元线性回归方法进行分析. 例如, 在含有两个自变量的一阶模型中添加一个交互作用项, 即

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon \tag{10.3}$$

若设 $x_3 = x_1 x_2$ 和 $\beta_3 = \beta_{12}$, 则 (10.3) 式可改写为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \tag{10.4}$$

这是一个有 3 个回归变量的标准多元线性回归模型. 回忆第 6~8 章, 我们为了定量表示二水平析因设计, 曾用类似于 (10.2) 式与 (10.4) 式的式子来表示经验模型. 再例如, 考虑 2 个变量的二阶**响应曲面模型**:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon \tag{10.5}$$

若设 $x_3 = x_1^2$, $x_4 = x_2^2$, $x_5 = x_1 x_2$, $\beta_3 = \beta_{11}$, $\beta_4 = \beta_{22}$, $\beta_5 = \beta_{12}$, 则上式改写为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon \tag{10.6}$$

这是一个线性回归模型. 在本书前面章节中, 我们已看到过这个模型. 一般来说, 任意一个对于参数 (β) 为线性的回归模型就叫做线性回归模型, 不论其生成的响应曲面形状如何.

本章将概述多元线性回归模型参数估计的方法. 这通常称为**模型拟合**. 我们还将对这些模型讨论假设检验和构造置信区间的方法, 以及检验模型拟合的适合性. 我们主要关注在设计过的实验 (designed experiment) 中有用的回归分析方法. 关于回归的更完整的论述, 可参阅 Montgomery, Peck, and Vining (2001) 以及 Myers (1990).

10.3 线性回归模型的参数估计

多元线性回归模型中, 通常用最小二乘法来估计回归系数. 设有 $n > k$ 个响应的观测值, 比方说, $y_1, y_2, \cdots, y_n$. 与每一个观测响应 y_i 对应, 每一个回归变量也有相应的观测值, 记 x_{ij} 表

示变量 x_j 的第 i 个观测值或水平. 数据如表 10.1 所示. 假定模型中误差项 ε 满足 $\mathrm{E}(\varepsilon)=0$, $\mathrm{V}(\varepsilon)=\sigma^2$, 且$\{\varepsilon_i\}$ 是不相关的随机变量.

表 10.1 多元线性回归的数据

y	x_1	x_2	$\cdots$	x_k
y_1	x_{11}	x_{12}	$\cdots$	x_{1k}
y_2	x_{21}	x_{22}	$\cdots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_n	x_{n1}	x_{n2}	$\cdots$	x_{nk}

可以用表 10.1 中列出的观测值写出 [(10.2) 式的] 模型

$$y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_kx_{ik}+\varepsilon_i=\beta_0+\sum_{j=1}^{k}\beta_jx_{ij}+\varepsilon_i,\quad i=1,2,\cdots,n \tag{10.7}$$

所谓最小二乘法, 就是选取 10.7 式中的 β 使得误差 ε_i 的平方和最小. 最小二乘函数是

$$L=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{k}\beta_jx_{ij}\right)^2 \tag{10.8}$$

将函数 L 关于 $\beta_0, \beta_1, \cdots, \beta_k$ 最小化. 最小二乘估计量 (记为 $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$) 必须满足

$$\frac{\partial L}{\partial \beta_0}|_{\hat{\beta}_0,\hat{\beta}_1,\cdots\hat{\beta}_k}=-2\sum_{i=1}^{n}\left(y_i-\hat{\beta}_0-\sum_{j=1}^{k}\hat{\beta}_jx_{ij}\right)=0 \tag{10.9a}$$

与

$$\frac{\partial L}{\partial \beta_j}|_{\hat{\beta}_0,\hat{\beta}_1,\cdots\hat{\beta}_k}=-2\sum_{i=1}^{n}\left(y_i-\hat{\beta}_0-\sum_{j=1}^{k}\hat{\beta}_jx_{ij}\right)x_{ij}=0\quad j=1,2,\cdots,k \tag{10.9b}$$

化简 (10.9) 式, 得

$$\begin{gathered}
n\hat{\beta}_0+\hat{\beta}_1\sum_{i=1}^{n}x_{i1}+\hat{\beta}_2\sum_{i=1}^{n}x_{i2}+\cdots+\hat{\beta}_k\sum_{i=1}^{n}x_{ik}=\sum_{i=1}^{n}y_i\\
\hat{\beta}_0\sum_{i=1}^{n}x_{i1}+\hat{\beta}_1\sum_{i=1}^{n}x_{i1}^2+\hat{\beta}_2\sum_{i=1}^{n}x_{i1}x_{i2}+\cdots+\hat{\beta}_k\sum_{i=1}^{n}x_{i1}x_{ik}=\sum_{i=1}^{n}x_{i1}y_i\\
\cdots\cdots\cdots\cdots\\
\hat{\beta}_0\sum_{i=1}^{n}x_{ik}+\hat{\beta}_1\sum_{i=1}^{n}x_{ik}x_{i1}+\hat{\beta}_2\sum_{i=1}^{n}x_{ik}x_{i2}+\cdots+\hat{\beta}_k\sum_{i=1}^{n}x_{ik}^2=\sum_{i=1}^{n}x_{ik}y_i
\end{gathered} \tag{10.10}$$

称这些方程为**最小二乘正规方程组** (least squares normal equations). 共有 $p=k+1$ 个正规方程, 正好和未知回归系数个数相对应. 正规方程组的解就是回归系数的最小二乘估计量 $\hat{\beta}_0$, $\hat{\beta}_1, \cdots, \hat{\beta}_k$.

如果用矩阵记号表示, 则求解正规方程组更为简便. 现在给出对应于 (10.10) 式的正规方程组的矩阵表示形式. 以观测值表示的模型, 即 (10.7) 式, 用矩阵记号可写为

$$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon}$$

其中

$$\boldsymbol{y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{X}=\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta}=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\varepsilon}=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

一般地, $\boldsymbol{y}$ 是一个 $(n\times1)$ 观测值向量, $\boldsymbol{X}$ 是自变量水平的一个 $(n\times p)$ 矩阵, $\boldsymbol{\beta}$ 是一个 $(p\times1)$ 回归系数向量, $\boldsymbol{\varepsilon}$ 是一个 $(n\times1)$ 随机误差向量.

要求得最小二乘估计向量 $\hat{\boldsymbol{\beta}}$, 只要使

$$L=\sum_{i=1}^{n}\varepsilon_i^2=\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}=(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})$$

取极小值. 注意 L 可表示为

$$L=\boldsymbol{y}'\boldsymbol{y}-\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y}-\boldsymbol{y}'\boldsymbol{X\beta}+\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}=\boldsymbol{y}'\boldsymbol{y}-2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y}+\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta} \tag{10.11}$$

这是因为 $\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y}$ 是一个 (1×1) 矩阵, 或标量, 所以它的转置 $(\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y})'=\boldsymbol{y}'\boldsymbol{X\beta}$ 是同一个标量. 最小二乘估计量必须满足

$$\frac{\partial L}{\partial \boldsymbol{\beta}}\Big|_{\hat{\boldsymbol{\beta}}}=-2\boldsymbol{X}'\boldsymbol{y}+2\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}=\boldsymbol{0}$$

简写为

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}=\boldsymbol{X}'\boldsymbol{y} \tag{10.12}$$

(10.12) 式是最小二乘正规方程组的矩阵形式, 它等同于 (10.10) 式. 为了求解正规方程组, 对 (10.12) 式两边乘以 $\boldsymbol{X}'\boldsymbol{X}$ 的逆. 于是, $\boldsymbol{\beta}$ 的最小二乘估计量是

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{10.13}$$

容易看出, 正规方程组的矩阵形式等同于它的标量形式. 详细写出 (10.12) 式, 得

$$\begin{bmatrix} n & \sum_{i=1}^{n}x_{i1} & \sum_{i=1}^{n}x_{i2} & \cdots & \sum_{i=1}^{n}x_{ik} \\ \sum_{i=1}^{n}x_{i1} & \sum_{i=1}^{n}x_{i1}^2 & \sum_{i=1}^{n}x_{i1}x_{i2} & \cdots & \sum_{i=1}^{n}x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n}x_{ik} & \sum_{i=1}^{n}x_{ik}x_{i1} & \sum_{i=1}^{n}x_{ik}x_{i2} & \cdots & \sum_{i=1}^{n}x_{ik}^2 \end{bmatrix}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}=\begin{bmatrix} \sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}x_{i1}y_i \\ \vdots \\ \sum_{i=1}^{n}x_{ik}y_i \end{bmatrix}$$

进行矩阵乘法, 就得到正规方程组 (即 (10.10) 式) 的数量形式. 易见, $\boldsymbol{X}'\boldsymbol{X}$ 是一个 $(p\times p)$ 对称矩阵, $\boldsymbol{X}'\boldsymbol{y}$ 是一个 $(p\times1)$ 列向量. 注意 $\boldsymbol{X}'\boldsymbol{X}$ 的特殊构造. $\boldsymbol{X}'\boldsymbol{X}$ 的对角线元素是 $\boldsymbol{X}$ 列元素的平方和, 对角线以外的元素是 $\boldsymbol{X}$ 列元素的叉积的和. 此外, $\boldsymbol{X}'\boldsymbol{y}$ 的元素是 $\boldsymbol{X}$ 列元素与观测值$\{y_i\}$ 的叉积的和.

所拟合的回归模型是

$$\hat{\boldsymbol{y}}=\boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{10.14}$$

其标量形式为

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_{ij}, \quad i = 1, 2, \cdots, n$$

实际观测值 y_i 与相应的拟合值 y_i 的差称为**残差** (residual), 记为 $e_i = y_i - \hat{y}_i$. $(n\times 1)$ 维残差向量记为

$$\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} \tag{10.15}$$

1. 估计 σ^2

我们常常需要估计 σ^2. 为了得到这个参数的估计量, 考虑残差平方和

$$SS_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = \boldsymbol{e}'\boldsymbol{e}$$

代入 $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$, 得

$$SS_E = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{y}'\boldsymbol{y} - 2\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$$

因为 $\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$, 上式化为

$$SS_E = \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} \tag{10.16}$$

称 (10.16) 式为**误差平方和或残差平方和**, 它有 $n-p$ 个自由度. 可以证明

$$\mathrm{E}(SS_E) = \sigma^2(n-p)$$

所以, σ^2 的一个无偏估计是

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} \tag{10.17}$$

2. 估计量的性质

由最小二乘法可以得到线性回归模型中参数 $\boldsymbol{\beta}$ 的无偏估计. 这只需按下面的方法对 $\hat{\boldsymbol{\beta}}$ 求期望就可以证明:

$$\begin{aligned}\mathrm{E}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}] = \mathrm{E}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] \\ &= \mathrm{E}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}] = \boldsymbol{\beta}\end{aligned}$$

这是因为 $\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}, (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}$. 于是, $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的无偏估计量.

$\hat{\boldsymbol{\beta}}$ 的方差特性用**协方差矩阵**表示为:

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) \equiv \mathrm{E}\{[\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})][\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})]'\} \tag{10.18}$$

它是一个对称矩阵, 它的第 i 个主对角线元素是单个回归系数 $\hat{\beta}_i$ 的方差, 它的第 (ij) 个元素是 $\hat{\beta}_i$ 与 $\hat{\beta}_j$ 之间的协方差. $\hat{\boldsymbol{\beta}}$ 的协方差矩阵是

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{10.19}$$

例 10.1 聚合物的黏度 (y) 和两个过程变量 —— 反应温度 (x_1) 与催化剂进料速率 (x_2)—— 的 16 个试验的观测值如表 10.2 所示. 我们将用这些数据拟合多元线性模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$\boldsymbol{X}$ 矩阵和 $\boldsymbol{y}$ 向量分别是

表 10.2 例 10.1 多元线性回归的数据

观测	温度 (x_1,℃)	催化剂进料速率 (x_2, lb/h)	黏度	观测	温度 (x_1, ℃)	催化剂进料速率 (x_2, lb/h)	黏度
1	80	8	2 256	9	94	12	2 364
2	93	9	2 340	10	93	11	2 379
3	100	10	2 426	11	97	13	2 440
4	82	12	2 293	12	95	11	2 364
5	90	11	2 330	13	100	8	2 404
6	99	8	2 368	14	85	12	2 317
7	81	8	2 250	15	86	9	2 309
8	96	10	2 409	16	87	12	2 328

$$
\boldsymbol{X}=\begin{bmatrix}1&80&8\\1&93&9\\1&100&10\\1&82&12\\1&90&11\\1&99&8\\1&81&8\\1&96&10\\1&94&12\\1&93&11\\1&97&13\\1&95&11\\1&100&8\\1&85&12\\1&86&9\\1&87&12\end{bmatrix},\qquad \boldsymbol{y}=\begin{bmatrix}2\,256\\2\,340\\2\,426\\2\,293\\2\,330\\2\,368\\2\,250\\2\,409\\2\,364\\2\,379\\2\,440\\2\,364\\2\,404\\2\,317\\2\,309\\2\,328\end{bmatrix}
$$

$\boldsymbol{X}'\boldsymbol{X}$ 矩阵是

$$
\boldsymbol{X}'\boldsymbol{X}=\begin{bmatrix}1&1&\cdots&1\\80&93&\cdots&87\\8&9&\cdots&12\end{bmatrix}\begin{bmatrix}1&80&8\\1&93&9\\\vdots&\vdots&\vdots\\1&87&12\end{bmatrix}=\begin{bmatrix}16&1\,458&164\\1\,458&133\,560&14\,946\\164&14\,946&1\,726\end{bmatrix}
$$

$\boldsymbol{X}'\boldsymbol{y}$ 向量是

$$
\boldsymbol{X}'\boldsymbol{y}=\begin{bmatrix}1&1&\cdots&1\\80&93&\cdots&87\\8&9&\cdots&12\end{bmatrix}\begin{bmatrix}2\,256\\2\,340\\\vdots\\2\,328\end{bmatrix}=\begin{bmatrix}37\,577\\3\,429\,550\\385\,562\end{bmatrix}
$$

$\boldsymbol{\beta}$ 的最小二乘估计是

$$
\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}
$$

或

$$
\hat{\boldsymbol{\beta}}=\begin{bmatrix}14.176\,004&-0.129\,746&-0.223\,453\\-0.129\,746&1.429\,184\times10^{-3}&-4.763\,947\times10^{-5}\\-0.223\,453&-4.763\,947\times10^{-5}&2.222\,381\times10^{-2}\end{bmatrix}\begin{bmatrix}37\,577\\3\,429\,550\\385\,562\end{bmatrix}
$$

$$= \begin{bmatrix} 1\,566.077\,77 \\ 7.621\,29 \\ 8.584\,85 \end{bmatrix}$$

回归系数保留两位小数时的最小二乘拟合是

$$\hat{y} = 1\,566.08 + 7.62x_1 + 8.58x_2$$

表 10.3 的前 3 列分别表示实际观测值 y_i、预测值或拟合值 $\hat{y}_i$ 和残差值. 图 10.1 是残差的正态概率图. 残差与预测值 $\hat{y}_i$ 的关系图和残差与两个变量 x_1 和 x_2 的关系图分别显示在图 10.2、图 10.3 和图 10.4 中. 正如在设计过的实验中一样, 残差图是回归模型中不可或缺的部分. 这些图形表明观测到的黏度值的方差随着黏度的增大而增加. 图 10.3 表明黏度的变异性随着温度的增加而增加.

表 10.3 例 10.1 的预测值、残差值及其他诊断值

观测 i	y_i	预测值 $\hat{y}_i$	残差 e_i	h_{ii}	学生化残差	D_i	学生化剔除残差
1	2 256	2 244.5	11.5	0.350	0.87	0.137	0.87
2	2 340	2 352.1	−12.1	0.102	−0.78	0.023	−0.77
3	2 426	2 414.1	11.9	0.177	0.80	0.046	0.79
4	2 293	2 294.0	−1.0	0.251	−0.07	0.001	−0.07
5	2 330	2 346.4	−16.4	0.077	−1.05	0.030	−1.05
6	2 368	2 389.3	−21.3	0.265	−1.52	0.277	−1.61
7	2 250	2 252.1	−2.1	0.319	−0.15	0.004	−0.15
8	2 409	2 383.6	25.4	0.098	1.64	0.097	1.76
9	2 364	2 385.5	−21.5	0.142	−1.42	0.111	−1.48
10	2 379	2 369.3	9.7	0.080	0.62	0.011	0.60
11	2 440	2 416.9	23.1	0.278	1.66	0.354	1.80
12	2 364	2 384.5	−20.5	0.096	−1.32	0.062	−1.36
13	2 404	2 396.9	7.1	0.289	0.52	0.036	0.50
14	2 317	2 316.9	0.1	0.185	0.01	0.000	<0.01
15	2 309	2 298.8	10.2	0.134	0.67	0.023	0.66
16	2 328	2 332.1	−4.1	0.156	−0.28	0.005	−0.27

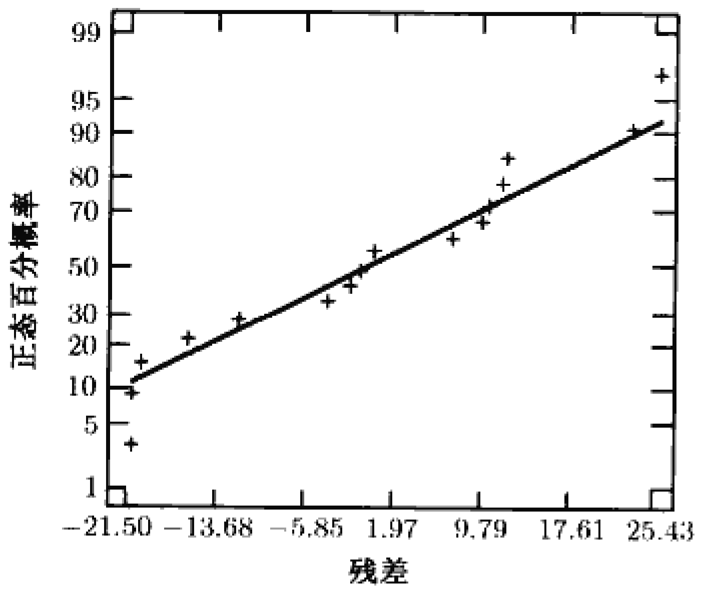


图 10.1 例 10.1 中残差的正态概率图

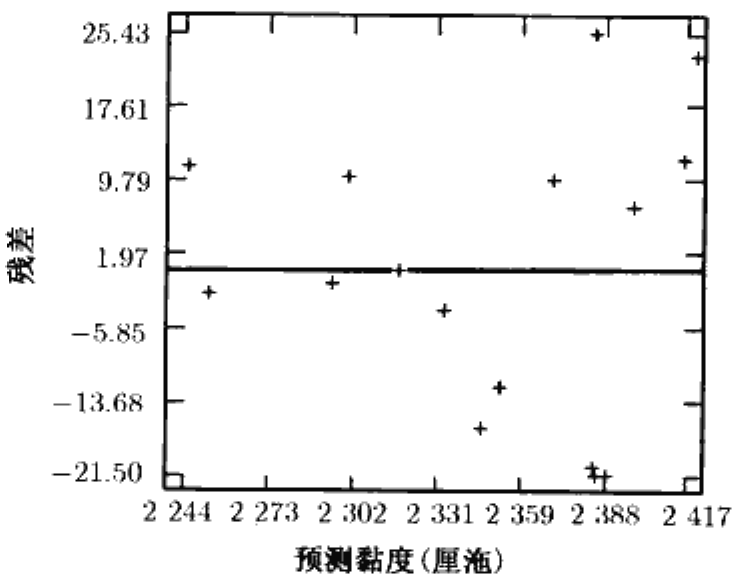


图 10.2 例 10.1 中残差与预测黏度的关系图

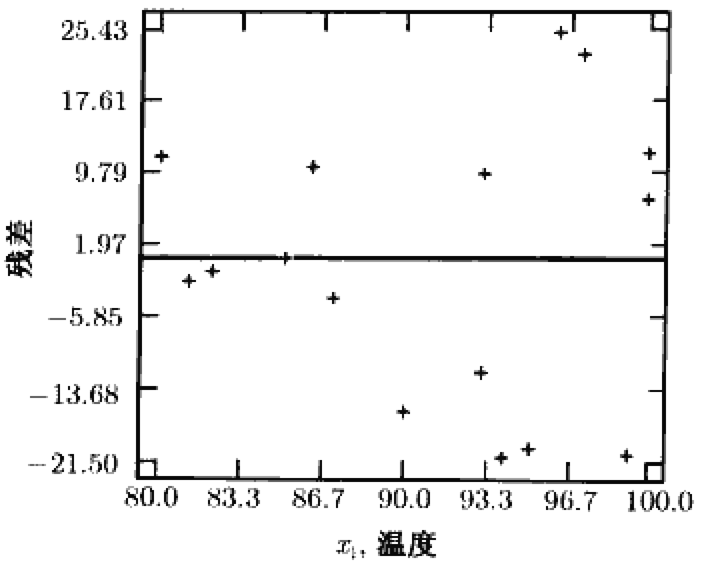


图 10.3　例 10.1 中残差与 x_1(温度) 的关系图

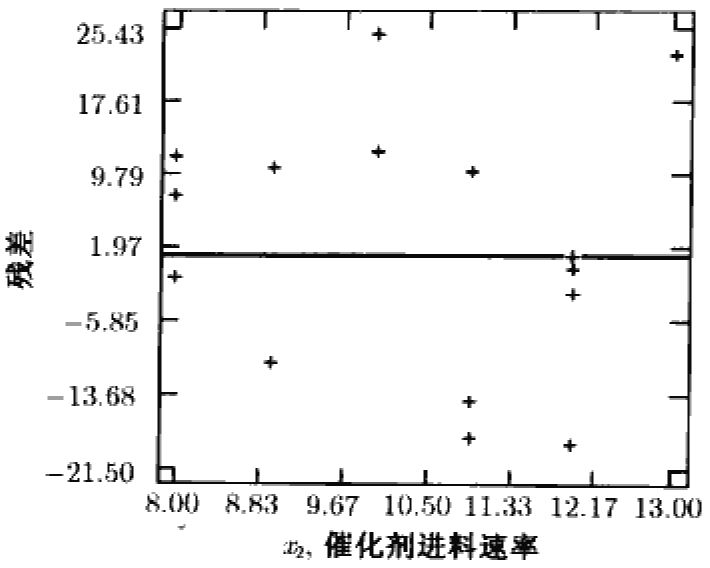


图 10.4　例 10.1 中残差与 x_2(催化剂进料速率) 的关系图

3. 使用计算机

几乎所有统计软件包都可以进行回归模型拟合. 表 10.4 列出了用 Minitab 软件拟合例 10.1 中黏度回归模型时所获得的部分输出结果. 在这个输出中有许多量我们已经熟悉了, 因为它们的含义类似于在设计过的实验中计算机输出的分析结果. 在本书前面的章节中, 我们已看到许多计算机输出结果, 在后面的章节中我们将详细讨论表 10.4 中方差分析和 t 检验的信息, 并将指出如何正确计算这些量.

表 10.4　例 10.1 黏度回归模型的 Minitab 输出

```
回归分析
The regression equation is
Viscosity = 1566 + 7.62Temp + 8.58 Feed Rate
Predictor              Coef       StDev        T         P
Constant            1566.08       61.59    25.43     0.000
Temp                 7.6213      0.6184    12.32     0.000
Feed Rat              8.585       2.439     3.52     0.004
S=16.36  R-Sq=92.7%  R-Sq(adj)=91.6%
Analysis of Variance
Source                   DF          SS       MS         F         P
Regression                2       44157    22079     82.50     0.000
Residual Error           13        3479      268
Total                    15       47636
Source      DF     Seq SS
Temp         1      40841
Feed Rat     1       3316
```

4. 在设计过的实验中拟合模型

我们常常用回归模型以定量的方式表示设计过的实验的结果. 下面我们给出一个完整的例子以说明如何来做. 在这个例子后面还有 3 个简单的例子, 用于说明在设计过的实验中回归分析的其他用途.

例 10.2　2^3 析因设计的回归分析

一位化学工程师研究一个过程的产率. 考察 3 个过程变量: 温度、压强以及催化剂浓度. 每个变量取低

与高两个水平, 工程师决定进行有 4 个中心点的 2^3 设计的实验. 设计与所得的产率如图 10.5 所示, 这里我们同时列出了设计因子的自然水平, 以及常在 2^k 析因设计中表示因子水平的 +1, −1 规范变量.

设工程师决定拟合只有主效应的模型, 即

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\varepsilon$$

对此模型, $\boldsymbol{X}$ 矩阵和 $\boldsymbol{y}$ 向量是

$$\boldsymbol{X}=\begin{bmatrix}1&-1&-1&-1\\1&1&-1&-1\\1&-1&1&-1\\1&1&1&-1\\1&-1&-1&1\\1&1&-1&1\\1&-1&1&1\\1&1&1&1\\1&0&0&0\\1&0&0&0\\1&0&0&0\\1&0&0&0\end{bmatrix},\quad \boldsymbol{y}=\begin{bmatrix}32\\46\\57\\65\\36\\48\\57\\68\\50\\44\\53\\56\end{bmatrix}$$

	过程变量			规范变量			
试验	温度 (℃)	压力 (psig)	浓度 (g/l)	x_1	x_2	x_3	产率, y
1	120	40	15	−1	−1	−1	32
2	160	40	15	1	−1	−1	46
3	120	80	15	−1	1	−1	57
4	160	80	15	1	1	−1	65
5	120	40	30	−1	−1	1	36
6	160	40	30	1	−1	1	48
7	120	80	30	−1	1	1	57
8	160	80	30	1	1	1	68
9	140	60	22.5	0	0	0	50
10	140	60	22.5	0	0	0	44
11	140	60	22.5	0	0	0	53
12	140	60	22.5	0	0	0	56

$x_1=\frac{温度-140}{20}$, $x_2=\frac{压强-60}{20}$, $x_1=\frac{浓度-22.5}{7.5}$

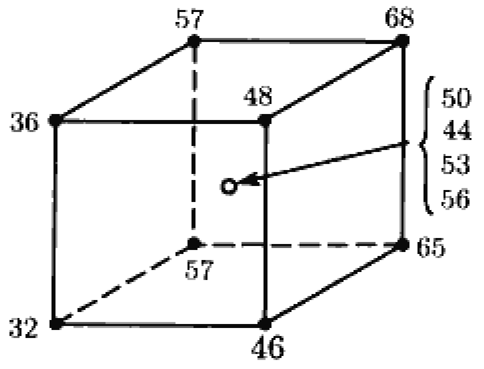


图 10.5 例 10.2 的实验设计

易得

$$\boldsymbol{X}'\boldsymbol{X}=\begin{bmatrix}12&0&0&0\\0&8&0&0\\0&0&8&0\\0&0&0&8\end{bmatrix},\quad \boldsymbol{X}'\boldsymbol{y}=\begin{bmatrix}612\\45\\85\\9\end{bmatrix}$$

因为 $\boldsymbol{X}'\boldsymbol{X}$ 是对角矩阵, 其逆也是对角的, 回归系数的最小二乘估计是

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}=\begin{bmatrix}1/12&0&0&0\\0&1/8&0&0\\0&0&1/8&0\\0&0&0&1/8\end{bmatrix}\begin{bmatrix}612\\45\\85\\9\end{bmatrix}=\begin{bmatrix}51.000\\5.625\\10.625\\1.125\end{bmatrix}$$

拟合的回归模型是

$$\hat{y}=51.000+5.625x_1+10.625x_2+1.125x_3$$

与我们在许多场合已使用过的相同，这些回归系数和通常的 2^3 设计分析法求得的效应估计量有密切的关系. 例如，温度的效应 (参阅图 10.5) 是

$$T = \bar{y}_{T+} - \bar{y}_{T-} = 56.75 - 45.50 = 11.25$$

注意 x_1 的回归系数是

$$(11.25)/2 = 5.625$$

也就是说，回归系数正好是通常效应估计的 1/2. 这一点对于 2^k 设计总是正确的. 如上所述，在第 6 章至第 8 章中，对好几个二水平实验，我们都用了这一结果来得到回归模型、拟合值以及残差. 此例表明，由 2^k 设计所得的效应估计量是最小二乘估计.

在例 10.2 中，因为 $\boldsymbol{X}'\boldsymbol{X}$ 是对角矩阵，逆矩阵容易求得. 这一点从直观上看有优越性，不仅因为计算简单，而且因为所有回归系数的估计量是不相关的，即，$\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) = 0$. 在收集数据之前，如果能够选择 x 变量的水平的话，我们就会设计实验使得 $\boldsymbol{X}'\boldsymbol{X}$ 为对角矩阵.

在实践中，这一点相对容易做到. 我们知道，$\boldsymbol{X}'\boldsymbol{X}$ 的对角线之外的元素是 $\boldsymbol{X}$ 列的叉积之和. 因此，必须使 $\boldsymbol{X}$ 的列间的内积等于零，也就是说，这些列必须是**正交的** (orthogonal). 为拟合回归模型而进行的实验设计，若具正交性，则称它为**正交设计**. 一般说来，2^k 析因设计是拟合多元线性回归模型的正交设计.

在设计过的实验中出现某些问题时，回归方法也是十分有用的. 这将在下面两个例子中给予说明.

例 10.3　有缺失观测值的 2^3 析因设计

考虑例 10.2 的有 4 个中心点的 2^3 析因设计. 假设完成实验后，缺失了各变量都是高水平的那个试验 (图 10.5 中的第 8 号试验). 这可能由多种原因造成，比如测量系统产生一个错误的读数，因子水平组合是不可能的搭配，实验单元可能有危险性，等等.

我们将用其余的 11 个数据拟合主效应模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

此时 $\boldsymbol{X}$ 矩阵和 $\boldsymbol{y}$ 向量是

$$\boldsymbol{X} = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 32 \\ 46 \\ 57 \\ 65 \\ 36 \\ 48 \\ 57 \\ 50 \\ 44 \\ 53 \\ 56 \end{bmatrix}$$

为估计模型参数，计算

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} 11 & -1 & -1 & -1 \\ -1 & 7 & -1 & -1 \\ -1 & -1 & 7 & -1 \\ -1 & -1 & -1 & 7 \end{bmatrix}, \quad \boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} 544 \\ -23 \\ 17 \\ -59 \end{bmatrix}$$

得

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

$$=\begin{bmatrix} 9.615\,38\times10^{-2} & 1.923\,07\times10^{-2} & 1.923\,07\times10^{-2} & 1.923\,07\times10^{-2} \\ 1.923\,07\times10^{-2} & 0.153\,85 & 2.884\,62\times10^{-2} & 2.884\,62\times10^{-2} \\ 1.923\,07\times10^{-2} & 2.884\,62\times10^{-2} & 0.153\,85 & 2.884\,62\times10^{-2} \\ 1.923\,07\times10^{-2} & 2.884\,62\times10^{-2} & 2.884\,62\times10^{-2} & 0.153\,85 \end{bmatrix}\begin{bmatrix} 544 \\ -23 \\ 17 \\ -59 \end{bmatrix}$$

$$=\begin{bmatrix} 51.25 \\ 5.75 \\ 10.75 \\ 1.25 \end{bmatrix}$$

因此, 拟合的模型是

$$\hat{y}=51.25+5.75x_1+10.75x_2+1.25x_3$$

将此模型与例 10.2 中由所有 12 个数据所得的模型比较. 回归系数非常相似. 因为回归系数与因子效应有密切的关系, 我们的结论并未受到缺失观测值的严重影响. 但是, 由于 $(\boldsymbol{X}'\boldsymbol{X})$ 与其逆不再是对角阵, 效应的估计量也不再正交.

例 10.4 设计因子水平不够精确

在进行一个设计过的实验时, 有时难以达到或保持设计所需的精确因子水平. 小的偏差并不重要, 但对大的偏差要多加关注. 在实验者不能得到所需因子水平的设计过的实验中, 回归方法很有用.

为了说明这个问题, 表 10.5 中的实验是例 10.2 的 2^3 设计的一个变形, 这里, 许多试验组合并不是设计所指定的真正水平的组合. 温度变量的控制似乎最为困难.

表 10.5 例 10.4 的实验设计

试验	过程变量			规范变量			产率, y
	温度 (℃)	压强 (psig)	浓度 (g/l)	x_1	x_2	x_3	
1	125	41	14	−0.75	−0.95	−1.133	32
2	158	40	15	0.90	−1	−1	46
3	121	82	15	−0.95	1.1	−1	57
4	160	80	15	1	1	−1	65
5	118	39	33	−1.10	−1.05	1.14	36
6	163	40	30	1.15	−1	1	48
7	122	80	30	−0.90	1	1	57
8	165	83	30	1.25	1.15	1	68
9	140	60	22.5	0	0	0	50
10	140	60	22.5	0	0	0	44
11	140	60	22.5	0	0	0	53
12	140	60	22.5	0	0	0	56

我们将拟合主效应模型

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\varepsilon$$

$\boldsymbol{X}$ 矩阵和 $\boldsymbol{y}$ 向量是

$$\boldsymbol{X}=\begin{bmatrix}1 & -0.75 & -0.95 & -1.133\\ 1 & 0.90 & -1 & -1\\ 1 & -0.95 & 1.1 & -1\\ 1 & 1 & 0 & -1\\ 1 & -1.10 & -1.05 & 1.4\\ 1 & 1.15 & -1 & 1\\ 1 & -0.90 & 1 & 1\\ 1 & 1.25 & 1.15 & 1\\ 1 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\end{bmatrix},\quad \boldsymbol{y}=\begin{bmatrix}32\\46\\57\\65\\36\\48\\57\\68\\50\\44\\53\\56\end{bmatrix}$$

为估计模型参数, 计算

$$\boldsymbol{X}'\boldsymbol{X}=\begin{bmatrix}12 & 0.60 & 0.25 & 0.267\,0\\ 0.60 & 8.18 & 0.31 & -0.140\,3\\ 0.25 & 0.31 & 8.537\,5 & -0.343\,7\\ 0.267\,0 & -0.140\,3 & -0.343\,7 & 9.243\,7\end{bmatrix},\quad \boldsymbol{X}'\boldsymbol{y}=\begin{bmatrix}612\\77.55\\100.7\\19.144\end{bmatrix}$$

得

$$\begin{aligned}\hat{\boldsymbol{\beta}}&=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\\ &=\begin{bmatrix}8.374\,47\times10^{-2} & -6.098\,71\times10^{-3} & -2.335\,42\times10^{-3} & -2.598\,33\times10^{-3}\\ -6.098\,71\times10^{-3} & 0.122\,89 & -4.207\,66\times10^{-3} & 1.884\,90\times10^{-3}\\ -2.335\,42\times10^{-3} & -4.207\,66\times10^{-3} & 0.117\,53 & 4.378\,51\times10^{-3}\\ -2.598\,33\times10^{-3} & 1.884\,90\times10^{-3} & 4.378\,51\times10^{-3} & 0.108\,45\end{bmatrix}\times\\ &\quad\begin{bmatrix}612\\77.55\\100.7\\19.144\end{bmatrix}=\begin{bmatrix}50.493\,91\\5.409\,96\\10.163\,16\\1.072\,45\end{bmatrix}\end{aligned}$$

回归系数保留两位小数, 拟合的模型是

$$\hat{y}=50.49+5.41x_1+10.16x_2+1.07x_3$$

将此模型与例 10.2 中的原模型比较, 原模型中因子水平是设计指定的真正水平, 两者之间只有非常小的差异. 这个实验结果的实际解释不会因为实验者不能精确达到所要的因子水平而受到严重影响.

例 10.5　在分式析因设计中, 分离有别名的交互作用

从第 8 章中可以看到, 用折叠法可以在一个分式析因设计中分离有别名的交互作用. 对一个分辨度为 III 的设计, 可以构造一个完全折叠设计, 其第二个分式设计是对原分式设计取反. 这个组合的设计可用于分离所有与两因子交互作用有别名的主效应.

使用完全折叠的难点在于, 它要求第二组试验的个数与原来的分式设计完全相同. 为了分离某些有别名的交互作用, 可以在原分式设计的基础上增加几个试验, 增加试验的个数少于全折叠所要求的. 部分折叠技术可以用于解决这个问题. 利用回归方法可以方便地看到部分折叠技术是如何起作用的, 有时甚至可以找到更有效的折叠设计.

例如, 假设我们要进行 2_{IV}^{4-1} 设计. 表 8-3 列出了这个设计的主分式设计, 在这个设计中, $I=ABCD$. 假设在获得前 8 个试验数据后, 最大的效应是 A, B, C, D(忽略别名是这些主效应的 3 因子交互作用) 和 $AB+CD$ 别名链. 虽然可以忽略另两个别名链, 但显然 AB 与 CD 中至少有一个偏大. 为发现哪个交互

作用是重要的, 我们当然可以做另一组替补的分式设计, 这需再做 8 个试验. 于是, 可用所有 16 个试验估计主效应和 2 因子交互作用. 另一种方法是进行包括 4 个附加试验的部分折叠设计.

可以用少于 4 个试验来分离别名 AB 和 CD. 假设我们希望拟合模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 + \beta_{34} x_3 x_4 + \varepsilon$$

其中 x_1, x_2, x_3, x_4 是表示 A, B, C, D 的规范变量. 用表 8-3 中的设计, 此模型的 $\boldsymbol{X}$ 矩阵是

$$\boldsymbol{X} = \begin{array}{c} \begin{matrix} \quad & x_1 & x_2 & x_3 & x_4 & x_1x_2 & x_3x_4 \end{matrix} \\ \begin{bmatrix} 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \end{array}$$

为了便于理解, 我们在各列的上方写出了变量名. 注意到 x_1x_2 列与 x_3x_4 列相同 (如所预期的, 因为 AB 或 x_1x_2 是 CD 或 x_3x_4 的别名), 这意味着 $\boldsymbol{X}$ 的各列之间不独立. 因此, 我们不能同时估计模型中的 β_{12} 和 β_{34}. 然而, 假设在替补的分式设计中把 $x_1 = -1$, $x_2 = -1$, $x_3 = -1$, $x_4 = 1$ 这个试验加入到原来的 8 个试验中, 则模型中的 $\boldsymbol{X}$ 矩阵变为

$$\boldsymbol{X} = \begin{array}{c} \begin{matrix} \quad & x_1 & x_2 & x_3 & x_4 & x_1x_2 & x_3x_4 \end{matrix} \\ \begin{bmatrix} 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & -1 \end{bmatrix} \end{array}$$

注意到, 现在列 x_1x_2 与 x_3x_4 不再相同, 这样我们就能够拟合同时含有交互作用 $x_1x_2(AB)$ 与 $x_3x_4(CD)$ 的回归模型. 回归系数的大小给出了交互作用重要性的信息.

尽管添加一个试验就可以分离交互作用 AB 和 CD, 但这种方法有一个缺点. 假设在前 8 个试验与稍后添加的试验之间存在一个时间效应 (或一个区组效应). 在 $\boldsymbol{X}$ 矩阵中加上区组列, 得

$$\boldsymbol{X} = \begin{array}{c} \begin{matrix} \quad & x_1 & x_2 & x_3 & x_4 & x_1x_2 & x_3x_4 & \text{区组} \end{matrix} \\ \begin{bmatrix} 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 \end{bmatrix} \end{array}$$

我们已假定在前 8 个试验中区组因子取低水平 "–", 在第 9 个试验中区组因子取高水平 "+". 可以看到, 每一列与区组列的叉积之和不等于零, 这意味着区组对处理不再是正交的, 或者说, 现在区组效应影响模型回归系数的估计. 为使区组正交化, 必须添加偶数次试验. 比如, 4 个试验

x_1	x_2	x_3	x_4
-1	-1	-1	1
1	-1	-1	-1
-1	1	1	1
1	1	1	-1

可以分离 AB 和 CD, 并得到正交区组 (用与前面类似的方法写出 $\boldsymbol{X}$ 矩阵, 就可以看到此结论). 由试验所需次数, 这等价于部分折叠.

通常直接检查由分式析因设计得到的简化模型的 $\boldsymbol{X}$ 矩阵, 并决定为了分离我们感兴趣的有别名的交互作用, 在原设计中应该增加哪些试验. 此外, 可以用本章稍后给出的回归模型的结果来评估特殊增大策略的影响. 还可以用基于计算机的方法构造用于分离有别名的效应的增大设计(参见第 8 章的补充材料). 第 11 章将讨论计算机生成的设计.

10.4 多元回归的假设检验

在多元线性回归问题中, 对模型参数的假设检验有助于度量模型的有效性. 本节将介绍几种重要的假设检验方法. 这些方法要求模型误差 ε_i 服从均值为零, 方差为 σ^2 的独立的正态分布, 简记为 $\varepsilon \sim \text{NID}(0, \sigma^2)$. 由此可知, 观测 y_i 服从均值为 $\beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}$, 方差为 σ^2 的独立的正态分布.

10.4.1 回归的显著性检验

回归方程的显著性检验用于判断响应变量 y 与回归变量 $x_1, x_2, \cdots, x_k$ 的子集之间是否存在线性关系. 相应的假设是

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0 \tag{10.20}$$

$$H_1: \beta_j \neq 0, \quad \text{至少有一个 } j$$

拒绝 (10.20) 式中的 H_0 意味着, 回归变量 $x_1, x_2, \cdots, x_k$ 中至少有一个变量对模型有显著的影响. 检验方法是方差分析法, 先将总平方和 SS_T 分解为模型 (或回归) 引起的平方和与残差 (或误差) 平方和, 即

$$SS_T = SS_R + SS_E \tag{10.21}$$

若零假设 H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ 为真, 则 SS_R/σ^2 服从 χ^2_k 分布, χ^2_k 的自由度等于模型中回归变量个数. 可以证明 SS_E/σ^2 服从 χ^2_{n-k-1} 分布且 SS_E 与 SS_R 相互独立. 对 H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ 的检验方法是计算

$$F_0 = \frac{SS_R/k}{SS_E/(n-k-1)} = \frac{MS_R}{MS_E} \tag{10.22}$$

若 F_0 超过 $F_{\alpha,k,n-k-1}$，则拒绝 H_0. 也可以用 P 值法进行假设检验，若统计量 F_0 的 P 值小于 α，则拒绝 H_0. 通常将检验过程汇总到如表 10.6 所示的方差分析表中.

表 10.6 多元回归的回归显著性方差分析

方差来源	平方和	自由度	均方	F_0
回归	SS_R	k	MS_R	MS_R/MS_E
误差或残差	SS_E	$n-k-1$	MS_E	
总和	SS_T	$n-1$		

容易得到 SS_R 的计算公式. 我们已在 (10.16) 式推出了 SS_E 的计算公式，即

$$SS_E = \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}$$

用 $SS_T = \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2/n = \boldsymbol{y}'\boldsymbol{y} - \left(\sum_{i=1}^{n} y_i\right)^2/n$ 改写上式，有

$$SS_E = \boldsymbol{y}'\boldsymbol{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} - \left[\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right]$$

又有

$$SS_E = SS_T - SS_R$$

因此，回归平方和为

$$SS_R = \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \tag{10.23}$$

误差平方和为

$$SS_E = \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} \tag{10.24}$$

总平方和为

$$SS_T = \boldsymbol{y}'\boldsymbol{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \tag{10.25}$$

几乎所有回归软件包都可以完成这些计算. 例如，表 10.4 就列出了用 Mintab 计算例 10.1 黏度回归模型所得的部分输出结果. 输出的上面部分是模型的方差分析表. 此例中回归的显著性检验的假设是

$$H_0: \beta_1 = \beta_2 = 0, \qquad H_1: \beta_j \neq 0,\ \text{至少有一个 } j$$

表 10.4 中，F 统计量 [(10.22) 式] 的 P 值非常小，所以，可以得到结论，两个变量 —— 温度 (x_1) 和进料速率 (x_2) —— 中至少有一个变量的回归系数不为零.

表 10.4 也列出了复决定系数 R^2 的值，这里的

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{10.26}$$

与设计过的实验中相同, R^2 用于度量由于在模型中使用回归变量 $x_1, x_2, \cdots, x_k$ 而造成的 y 变异性的减少量. 然而, 正如我们前面已注意到的, 具有大的 R^2 值的回归模型并不意味着它一定是好的模型. 在模型中增加一个变量, 不管它是否统计显著, 总会增加 R^2 的值. 于是, 可能会发生这样的情况, 一个模型有大的 R^2 值, 而用它对新的观测进行预测或求响应期望的估计时效果却很差.

因为向模型中添加新的项时 R^2 总是增加的, 有些建立回归模型的人更愿意用**调整的R^2统计量**, 其定义为

$$R^2_{\text{adj}} = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} = 1 - \left(\frac{n-1}{n-p}\right)(1-R^2) \tag{10.27}$$

一般而言, 当模型中增加变量时, 调整的 R^2 统计量并不总是增加的. 实际上, 如果增加了不必要的项, R^2_{adj} 的值就会减少.

以黏度回归模型为例. 模型的调整 R^2 在表 10.4 中已列出, 其计算过程为

$$R^2_{\text{adj}} = 1 - \left(\frac{n-1}{n-p}\right)(1-R^2) = 1 - \left(\frac{15}{13}\right)(1-0.926\ 97) = 0.915\ 735$$

它非常接近普通的 R^2. 如果 R^2 与 R^2_{adj} 有极大的不同, 模型中很可能含有不显著的项.

10.4.2 回归系数的个别检验和分组检验

我们经常感兴趣于检验各回归系数的假设. 这类检验可用于判定回归模型中每个回归变量的重要性. 例如, 在模型中添加一些变量或者删去模型的一个或多个变量, 模型可能会更为有效.

向回归模型添加变量总会导致回归平方和增加与误差平方和减少. 我们必须确定, 用增加模型变量的方法使回归平方和增加是否值得. 此外, 向模型中添加一个不重要的变量, 实际上会增大误差的均方, 因而会使模型的效用下降.

检验任意一个回归系数的显著性, 比如 β_j, 假设是

$$H_0: \beta_j = 0, \qquad H_1: \beta_j \neq 0$$

若不拒绝 H_0: $\beta_j= 0$, 则表示可以从模型中删去变量 x_j. 这个假设的检验统计量是

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \tag{10.28}$$

这里, C_{jj} 是 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 对应于 $\hat{\beta}_j$ 的对角元素. 当 $|t_0| > t_{\alpha/2,n-k-1}$ 则拒绝原假设 H_0:$\beta_j= 0$. 注意, 这实际上是偏检验或边缘检验, 因为回归系数 $\hat{\beta}_j$ 依赖于模型中所有其他变量 $x_i(i \neq j)$.

通常称 (10.28) 式中的分母 $\sqrt{\hat{\sigma}^2 C_{jj}}$ 为回归系数 $\hat{\beta}_j$ 的**标准误**, 即

$$se(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}} \tag{10.29}$$

因此, (10.28) 式中的检验统计量可等价地写为

$$t_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \tag{10.30}$$

大多数回归的计算机程序对每个模型参数给出 t 检验. 例如, 考虑表 10.4, 它是例 10.1 的 Mintab 输出. 输出的上面部分给出了每个参数的最小二乘估计、标准误、t 统计量以及对应的 P 值. 对此模型, 我们的结论是两个变量 (温度和进料速率) 都是显著的.

也可以直接检验某个特殊变量, 如 x_j, 在给定模型中的其他变量 x_i $(i \neq j)$ 时, 对于回归平方和的贡献. 检验方法就是普通的回归显著性检验, 被称为**附加平方和法**. 这种方法也用于研究回归变量的子集对模型的贡献. 考虑有 k 个变量的回归模型:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

其中 $\boldsymbol{y}$ 是 $(n\times 1)$ 向量, $\boldsymbol{X}$ 是 $(n \times p)$ 矩阵, $\boldsymbol{\beta}$ 是 $(p\times 1)$ 向量, $\boldsymbol{\varepsilon}$ 是 $(n\times 1)$ 向量, 且 $p = k + 1$. 我们想判断回归变量的子集 $x_1, x_2, \cdots, x_r$ ($r < k$) 是否对回归模型有显著性贡献. 将回归系数向量划分为

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}$$

其中 $\boldsymbol{\beta}_1$ 是 $(r\times 1)$ 向量, $\boldsymbol{\beta}_2$ 是 $[(p-r)\times 1]$ 向量. 要检验假设

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{0}, \qquad H_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{0} \tag{10.31}$$

模型可以写为

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \tag{10.32}$$

其中 $\boldsymbol{X}_1$ 表示 $\boldsymbol{X}$ 中和 $\boldsymbol{\beta}_1$ 有关的列, $\boldsymbol{X}_2$ 表示 $\boldsymbol{X}$ 中和 $\boldsymbol{\beta}_2$ 有关的列.

对**全模型** (包括 $\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$), 已知 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, 包含截距在内的所有变量的回归平方和为

$$SS_R(\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} \quad (p \text{ 个自由度})$$

且

$$MS_E = \frac{\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{y}}{n-p}$$

称 $SS_R(\boldsymbol{\beta})$ 为 $\boldsymbol{\beta}$ 的回归平方和. 为求出 $\boldsymbol{\beta}_1$ 对回归的贡献, 我们在假定零假设 H_0: $\boldsymbol{\beta}_1 = \boldsymbol{0}$ 为真时来拟合模型. 由 (10.32) 式得 $\boldsymbol{\beta}_1 = \boldsymbol{0}$ 时的**简化模型**:

$$\boldsymbol{y} = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \tag{10.33}$$

$\boldsymbol{\beta}_2$ 的最小二乘估计是 $\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{y}$, 且

$$SS_R(\boldsymbol{\beta}_2) = \hat{\boldsymbol{\beta}}'\boldsymbol{X}_2'\boldsymbol{y} \quad (p-r \text{ 个自由度}) \tag{10.34}$$

在模型中已有 $\boldsymbol{\beta}_2$ 出现的条件下, $\boldsymbol{\beta}_1$ 的回归平方和是

$$SS_R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta}_2) \tag{10.35}$$

此平方和有 r 个自由度, 称为 $\boldsymbol{\beta}_1$ 的 "附加平方和". 注意 $SS_R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2)$ 是由模型增加了变量 x_1, $x_2, \cdots, x_r$ 而引起的回归平方和的增加量.

现在, $SS_R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2)$ 与 MS_E 相互独立, 而零假设 $\boldsymbol{\beta}_1=0$ 可以用统计量

$$F_0=\frac{SS_R(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2)/r}{MS_E} \tag{10.36}$$

来检验. 当 $F_0>F_{\alpha,r,n-p}$ 时拒绝 H_0, 结论为, $\boldsymbol{\beta}_1$ 中至少有一个参数不为零, 从而在 $\boldsymbol{X}_1$ 内的变量 $x_1, x_2, \cdots, x_r$ 中至少有一个变量对回归模型有显著影响. 有人称 (10.36) 式的检验为偏 F 检验.

偏 F 检验是非常有用的. 可以通过计算

$$SS_R(\beta_j|\beta_0,\beta_1,\cdots,\beta_{j-1},\beta_{j+1},\cdots,\beta_k)$$

来度量最后一个添加到模型中去的变量 x_j 的贡献. 这是向包含了 $x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_k$ 的模型中添加 x_j 而引起的回归平方和的增加量. 注意对单一变量 x_j 的偏 F 检验等价于 (10.28) 式中的 t 检验. 因此, 偏 F 检验是我们度量变量集合的效应的更一般化的方法.

例 10.6 考虑例 10.1 中的黏度数据. 假定要研究变量 x_2(进料速率) 对模型的作用. 即所要检验的假设是

$$H_0: \beta_2=0, \qquad H_1: \beta_2\neq 0$$

这需要求出 β_2 的附加平方和

$$SS_R(\beta_2|\beta_1,\beta_0)=SS_R(\beta_0,\beta_1,\beta_2)-SS_R(\beta_0,\beta_1)=SS_R(\beta_1,\beta_2|\beta_0)-SS_R(\beta_1|\beta_0)$$

由检验了回归显著性的表 10.4, 我们有

$$SS_R(\beta_1,\beta_2|\beta_0)=44\ 157.1$$

在表中它被称为模型平方和. 这个平方和有 2 个自由度.

简化模型是

$$y=\beta_0+\beta_1x_1+\varepsilon$$

此模型的最小二乘拟合为

$$\hat{y}=1\ 652.395\ 5+7.639\ 7x_1$$

其 (自由度为 1 的) 回归平方和是

$$SS_R(\beta_1|\beta_0)=40\ 840.8$$

注意, 这个 $SS_R(\beta_1|\beta_0)$ 显示在表 10.4 中 Mintab 输出的底部的 "Seq SS" 项中. 因此,

$$SS_R(\beta_2|\beta_0,\beta_1)=44\ 157.1-40\ 840.8=3\ 316.3$$

其自由度为 $2-1=1$. 这是向已包含了 x_1 的模型中添加 x_2 而引起的回归平方和的增加量, 已显示在表 10.4 中 Mintab 输出的底部. 为了检验 H_0: $\beta_2=0$, 由检验统计量可得

$$F_0=\frac{SS_R(\beta_2|\beta_0,\beta_1)/1}{MS_E}=\frac{3\ 316.3/1}{267.604}=12.392\ 6$$

注意, F_0 的分母为全模型中的 MS_E (表 10.4). 由 $F_{0.05,1,13}=4.67$, 我们拒绝 H_0: $\beta_2=0$, 并认为 x_2 (进料速率) 对模型有显著影响.

这个偏 F 检验仅涉及单一回归变量, 此时它等价于 t 检验, 因为自由度为 v 的 t 统计量的平方就是自由度为 1 和 v 的 F 统计量. 为了理解这一点, 查看表 10.4 中对 H_0: $\beta_2=0$ 检验的 t 统计量 $t_0=3.520\ 3$, 因此, $t_0^2=(3.520\ 3)^2=12.392\ 5\approx F_0$.

10.5 多元回归的置信区间

我们常常要对回归系数$\{\beta_j\}$ 以及回归模型中其他感兴趣的量构造置信区间估计. 这些置信区间的推导过程需要假定误差$\{\varepsilon_i\}$ 服从均值为零、方差为 σ^2 的独立的正态分布, 与 10.4 节的假设检验中的假定相同.

10.5.1 单个回归系数的置信区间

因为最小二乘估计量 $\hat{\boldsymbol{\beta}}$ 是观测的线性组合, 所以, $\hat{\boldsymbol{\beta}}$ 服从均值为 $\boldsymbol{\beta}$、协方差矩阵为 $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 的正态分布. 故统计量

$$\frac{\hat{\beta}_j-\beta_j}{\sqrt{\hat{\sigma}^2C_{jj}}},\quad j=0,1,\cdots,k \tag{10.37}$$

服从自由度为 $n-p$ 的 t 分布, 这里, C_{jj} 是矩阵 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 的第 (jj) 个元素, $\hat{\sigma}^2$ 是由 (10.17) 式得到的误差方差的估计. 因此, 回归系数 β_j ($j=0,1,\cdots,k$) 的 100(1 − α)%置信区间是

$$\hat{\beta}_j-t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2C_{jj}}\leqslant\beta_j\leqslant\hat{\beta}_j+t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2C_{jj}} \tag{10.38}$$

因为 $se(\hat{\beta}_j)=\sqrt{\hat{\sigma}^2C_{jj}}$, 这个置信区间也可写成

$$\hat{\beta}_j-t_{\alpha/2,n-p}se(\hat{\beta}_j)\leqslant\hat{\beta}_j\leqslant\hat{\beta}_j+t_{\alpha/2,n-p}se(\hat{\beta}_j)$$

例 10.7 求例 10.1 中 β_1 的 95%置信区间. 现在 $\hat{\beta}_1=7.621\ 29$, 由 $\hat{\sigma}^2=267.604$, $C_{11}=1.429\ 184\times10^{-3}$, 可得

$$\hat{\beta}_1-t_{0.025\ 13}\sqrt{\hat{\sigma}^2C_{11}}\leqslant\beta_1\leqslant\hat{\beta}_1+t_{0.025\ 13}\sqrt{\hat{\sigma}^2C_{11}}$$

$$7.621\ 29-2.16\sqrt{(267.604)(1.429\ 184\times10^{-3})}\leqslant$$
$$\beta_1\leqslant7.621\ 29+2.16\sqrt{(267.604)(1.429\ 184\times10^{-3})}$$

$$7.621\ 29-2.16(0.618\ 4)\leqslant\beta_1\leqslant7.621\ 29+2.16(0.618\ 4)$$

故 β_1 的 95%置信区间是

$$6.285\ 5\leqslant\beta_1\leqslant8.957\ 0$$

10.5.2 平均响应的置信区间

我们还可以得到在特定点 (比如 x_{01}, x_{02}, $\cdots$, x_{0k}) 上响应均值的置信区间. 首先, 设向量

$$\boldsymbol{x}_0=\begin{bmatrix}1\\x_{01}\\x_{02}\\\vdots\\x_{0k}\end{bmatrix}$$

在该点的平均响应为

$$\mu_{y|\boldsymbol{x}_0}=\beta_0+\beta_1x_{01}+\beta_2x_{02}+\cdots+\beta_kx_{0k}=\boldsymbol{x}_0'\boldsymbol{\beta}$$

在该点的平均响应的估计为

$$\hat{y}(\boldsymbol{x}_0)=\boldsymbol{x}_0'\hat{\boldsymbol{\beta}} \tag{10.39}$$

因为 $\mathrm{E}[\hat{y}(\boldsymbol{x}_0)]=\mathrm{E}(\boldsymbol{x}_0'\hat{\boldsymbol{\beta}})=\boldsymbol{x}_0'\boldsymbol{\beta}=\boldsymbol{\mu}_{y|\boldsymbol{x}_0}$, 所以这个估计量是无偏的. $\hat{y}(\boldsymbol{x}_0)$ 的方差为

$$\mathrm{V}[\hat{y}(\boldsymbol{x}_0)]=\sigma^2\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0 \tag{10.40}$$

因此, 在点 $x_{01}, x_{02}, \cdots, x_{0k}$ 处的平均响应的 $100(1-\alpha)\%$置信区间是

$$\hat{y}(\boldsymbol{x}_0)-t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0}\leqslant\mu_{y|\boldsymbol{x}0}\leqslant\hat{y}(\boldsymbol{x}_0)+t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0} \tag{10.41}$$

10.6　新响应观测的预测

回归模型可用于预测响应 y 关于回归变量的特定值 (如 $x_{01}, x_{02}, \cdots, x_{0k}$) 的观测. 设 $\boldsymbol{x}_0'=[1, x_{01}, x_{02}, \cdots, x_{0k}]$, 那么, 在点 $x_{01}, x_{02}, \cdots, x_{0k}$ 处观测 y_0 的点估计可用 (10.39) 式

$$\hat{y}(\boldsymbol{x}_0)=\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}$$

计算出来. y_0 的 $100(1-\alpha)\%$**预测区间** (prediction interval) 是

$$\begin{aligned}&\hat{y}(\boldsymbol{x}_0)-t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1+\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0)}\leqslant\\&y_0\leqslant\hat{y}(\boldsymbol{x}_0)+t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1+\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0)}\end{aligned} \tag{10.42}$$

在点 $x_{01}, x_{02}, \cdots, x_{0k}$ 处预测新观测和估计平均响应时, 对于位于包含原有观测范围之外的情形, 我们必须小心谨慎地推断. 这是因为很可能出现这样的情况: 模型在原有的数据范围内拟合得很好, 但在该范围之外就不能很好地拟合了.

(10.42) 式中的预测区间有许多应用. 其中之一是用于析因设计或分式析因设计中的确认实验. 在一个确认实验中, 我们要检验从原实验中得到的模型, 以判断我们的解释是否正确. 通常, 先用模型来预测设计空间内一些感兴趣的点上的响应, 然后比较预测的响应与在那些点上另行试验所获得的实际观测. 我们已在第 8 章中用例 8.1 的 2^{4-1} 分式析因设计说明了这种方法. 确认的一种有用的方法是看新的观测是否落在该点预测区间的范围内.

为了说明这一点, 仍考虑例 8.1. 这个实验的解释表明, 4 个主效应中的 3 个 (A, C, D) 以及 2 因子交互效应中的 2 个 (AC 和 AD) 是重要的效应. 用 A, B, D 因子在高水平、C 因子在低水平的点作为合理的确认试验的试验点, 该点的响应的预测值是 100.25. 如果分式析因实验已被正确解释且响应模型有效, 我们期望该点的观测值落在由 (10.42) 式计算的预测区间内. 这个区间是容易计算的. 因为 2^{4-1} 是正交设计且模型包含 6 项 (截距, 3 个主效应, 2 个二因子交互效应), 矩阵 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 有特别简单的形式, 即 $(\boldsymbol{X}'\boldsymbol{X})^{-1}=\frac{1}{8}\boldsymbol{I}_6$. 所考虑的点的坐标是 $x_1=1, x_2=1, x_3=-1, x_4=1$, 但因为 B (即 x_2) 不在模型中, 而 2 个交互效应 AC 和 AD (x_1x_3 和 x_1x_4) 在模型中, 点 $\boldsymbol{x}_0$ 的坐标可记为 $\boldsymbol{x}_0'=[1, x_1, x_3, x_4, x_1x_3, x_1x_4]=[1,1,-1,1,-1,1]$. 易知此模型 (有 2 个自由度的) σ^2 的估计是 $\hat{\sigma}^2=3.25$. 由 (10.42) 式, 在此点的观测的 95%预测区间是

$$\hat{y}(\boldsymbol{x}_0)-t_{0.025,2}\sqrt{\hat{\sigma}^2(1+\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0)}\leqslant y_0\leqslant\hat{y}(\boldsymbol{x}_0)+t_{0.025,2}\sqrt{\hat{\sigma}^2(1+\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0)}$$

$$100.25 - 4.30\sqrt{3.25\left(1 + \boldsymbol{x}_0'\frac{1}{8}\boldsymbol{I}_6\boldsymbol{x}_0\right)} \leqslant y_0 \leqslant 100.25 + 4.30\sqrt{3.25\left(1 + \boldsymbol{x}_0'\frac{1}{8}\boldsymbol{I}_6\boldsymbol{x}_0\right)}$$

$$100.25 - 4.30\sqrt{3.25(1+0.75)} \leqslant y_0 \leqslant 100.25 + 4.30\sqrt{3.25(1+0.75)}$$

$$100.25 - 10.25 \leqslant y_0 \leqslant 100.25 + 10.25$$

$$90 \leqslant y_0 \leqslant 110.50$$

因此, 我们期望对 A, B, D 因子在高水平、C 因子在低水平的点上进行确认试验的响应变量渗透率的观测应落在 90 至 110.50 之间. 实际观测值是 104. 成功的确认试验对正确解释分式析因实验提供了一些保障.

10.7 回归模型的诊断

正如我们所强调的, 在设计过的实验中, **模型适合性检验**是数据分析过程的一个重要部分. 它与建立回归模型同样重要, 如例 10.1, 我们总是用设计过的实验中的**残差图**来检查回归模型. 一般需要进行: (1) 检查所拟合的模型, 以保证它对真实模型提供一个足够精度的近似; (2) 验证所有最小二乘回归的假定都满足. 如若回归模型不是一个适当的拟合, 它会给出一个拙劣的或起误导作用的结果.

除了残差图, 回归中还常用其他一些模型诊断方法. 本节将简要地叙述这类方法中的一部分. 更完整的叙述见 Montgomery, Peck, and Vining (2001) 以及 Myers (1990).

10.7.1 尺度残差和 PRESS

1. 标准化残差与学生化残差

许多建立模型的人更愿意用对应的**尺度残差**而不愿用普通最小二乘残差. 这些尺度残差常常比普通残差表达更多的信息.

一种尺度残差是**标准化残差**:

$$d_i = \frac{e_i}{\hat{\sigma}} \quad i = 1, 2, \cdots, n \tag{10.43}$$

其中的 $\hat{\sigma}$ 一般用 $\hat{\sigma} = \sqrt{MS_E}$ 计算. 标准化残差的均值为零, 方差近似为 1, 因此, 通常可用于查找**异常值**. 大部分的标准化残差应落在区间 $-3 \leqslant d_i \leqslant 3$ 内, 任意一个标准化残差在此范围之外的观测相对于观测响应而言很可能是异常观测. 应该仔细检查这些异常点, 因为它们可能表示出现了简单的数据录入错误, 也可能表示出现了更严重的问题, 如在回归变量所在的区域中, 拟合模型逼近真实响应曲面严重不足.

(10.43) 式的标准化过程就是用残差除以它的近似平均标准差来对残差尺度化. 在某些数据集合中, 各残差之间的标准差有很大的不同. 现在我们提出一种考虑此问题的尺度化方法.

对应于观测值 y_i 的拟合值 $\hat{y}_i$ 的向量是

$$\hat{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{H}\boldsymbol{y} \tag{10.44}$$

$n \times n$ 矩阵 $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ 通常称为 "帽子" 矩阵, 因为它在观测值向量上画出符号变成拟合值向量. 帽子矩阵及其性质在回归分析中扮演着重要的角色.

拟合模型的残差可以用矩阵符号方便地写成

$$\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}$$

残差的协方差矩阵是

$$\mathrm{Cov}(\boldsymbol{e}) = \sigma^2(\boldsymbol{I} - \boldsymbol{H}) \tag{10.45}$$

矩阵 $\boldsymbol{I} - \boldsymbol{H}$ 一般不是对角阵, 所以各残差的方差不相等且它们是相关的.

于是, 第 i 个残差的方差是

$$\mathrm{V}(e_i) = \sigma^2(1 - h_{ii}) \tag{10.46}$$

这里 h_{ii} 是 $\boldsymbol{H}$ 的第 i 个对角元素. 由于 $0 \leqslant h_{ii} \leqslant 1$, 所以用残差均方和 MS_E 来估计残差的方差实际上高估了 $\mathrm{V}(e_i)$. 由于 h_{ii} 是 x 空间中第 i 点位置的度量, 所以 e_i 的方差依赖于点 x_i 所处的位置. 一般地, x 空间中心点附近残差的方差比远离中心点的残差的方差大. 在远离中心点更可能发生违背模型假定的情况, 这些违背假定的情形可能难以用检测 e_i (或 d_i) 发现, 因为它们的残差常常很小.

我们推荐用不相等的方差来尺度化残差. 建议画出**学生化残差**

$$r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}} \quad i = 1, 2, \cdots, n \tag{10.47}$$

的图形来代替 e_i (或 d_i) 的图形, 这里的 $\hat{\sigma}^2 = MS_E$. 当模型形式正确时, 不论 x_i 位于何处, 学生化残差的方差都是常数 $\mathrm{V}(r_i) = 1$. 在许多情况下, 残差的方差是稳定的, 特别是对于大的数据集. 在这种情况下, 标准化残差与学生化残差之间差别很小. 此时标准化残差和学生化残差表达了同样的信息. 然而, 因为任一残差和 h_{ii} 都大的点极可能影响最小二乘拟合, 所以我们通常建议察看一下学生化残差. 表 10.3 列出了例 10.1 黏度回归模型的帽子矩阵对角元 h_{ii} 和学生化残差.

2. PRESS 残差

预测误差平方和 (PRESS) 提供了一种有用的残差尺度. 为了计算 PRESS, 先选择一个观测, 比如第 i 个观测. 然后用其余 $n-1$ 个观测拟合模型, 并用这个回归方程预测保留的观测 y_i. 记这个预测值为 $\hat{y}_{(i)}$, 得到第 i 点的预测误差 $e_{(i)} = y_i - \hat{y}_{(i)}$. 通常称此预测误差为第 i 个 PRESS 残差. 对每一个观测 $i = 1, 2, \cdots, n$ 重复此过程, 可以得到 n 个 PRESS 残差 $e_{(1)}, e_{(2)}, \cdots, e_{(n)}$ 的集合. 于是, 定义 PRESS 统计量为 n 个 PRESS 残差的平方和:

$$\mathrm{PRESS} = \sum_{i=1}^{n} e_{(i)}^2 = \sum_{i=1}^{n} [y_i - \hat{y}_{(i)}]^2 \tag{10.48}$$

PRESS 用每一个有 $n-1$ 个观测的子集作为估计数据集, 又用每个观测轮流作为预测数据集.

初看起来, 计算 PRESS 需要拟合 n 个不同的回归方程. 然而, PRESS 可以用关于所有 n 个观测拟合的单一回归方程的结果计算得到. 可以证明, 第 i 个 PRESS 残差是

$$e_{(i)} = \frac{e_i}{1 - h_{ii}} \tag{10.49}$$

因为 PRESS 正是 PRESS 残差的平方和, 所以其计算公式可写为

$$\mathrm{PRESS} = \sum_{i=1}^{n} \left(\frac{e_i}{1 - h_{ii}}\right)^2 \tag{10.50}$$

由 (10.49) 式可以看到, PRESS 残差是对普通残差按照帽子矩阵对角元 h_{ii} 加权得出的结果. 对 h_{ii} 较大的数据点, PRESS 残差就较大. 这些观测一般将成为强影响点. 普通残差和 PRESS 残差相差很大的点意味着, 虽然模型对数据拟合得很好, 但是如果没有该点则所建模型预测会很差. 10.7.2 节将讨论影响的一些其他度量.

最后, 可以用 PRESS 近似计算预测 R^2, 如

$$R^2_{\text{预测}} = 1 - \frac{\text{PRESS}}{SS_T} \tag{10.51}$$

这个统计量给出了回归方程预测能力的一个指标. 例 10.1 黏度回归模型中, 可以用普通残差以及表 10.3 中给出的 h_{ii} 的值来计算 PRESS 残差. 相应的 PRESS 统计量的值是 PRESS = 5 207.7, 于是

$$R^2_{\text{预测}} = 1 - \frac{\text{PRESS}}{SS_T} = 1 - \frac{5\ 207.7}{47\ 635.9} = 0.890\ 7$$

因此, 我们能够期待, 在预测新观测时, 这个模型可 "解释" 约 89%的变异性, 而最小二乘拟合可解释原数据的约 93%的变异性. 按此标准, 这个模型总的预测能力似乎是非常令人满意的.

3. **学生化剔除残差** (R-Student)

上面讨论的学生化残差 r_i 常用于异常点诊断. 在计算 r_i 时, 习惯上以 MS_E 作为 σ^2 的一个估计. 因为 MS_E 是用所有 n 个观测拟合模型而得到的内部产生的 σ^2 估计, 这意味着它是残差的内部的尺度. 另一种基于剔除第 i 个观测的数据集的方法也可以用作 σ^2 的估计. 记这样的 σ^2 的估计为 $S^2_{(i)}$. 可以证明

$$S^2_{(i)} = \frac{(n-p)MS_E - e_i^2/(1-h_{ii})}{n-p-1} \tag{10.52}$$

(10.52) 式中的 σ^2 的估计用于取代 MS_E, 产生一个外部的学生化残差, 称为学生化剔除残差, 计算公式为

$$t_i = \frac{e_i}{\sqrt{S^2_{(i)}(1-h_{ii})}} \quad i = 1, 2, \cdots, n \tag{10.53}$$

在大多数情况下, t_i 与学生化残差 r_i 只有极小的差异. 然而, 如果第 i 个观测影响较大, 那么 $S^2_{(i)}$ 与 MS_E 会有显著不同, 则学生化剔除残差对该点更为敏感. 因此, 在通常的假定下, t_i 的分布为 t_{n-p-1}. 故学生化剔除残差为探测异常点提供了一种可由假设检验进行的更正规的方法. 表 10.3 列出了例 10.1 黏度的回归模型的学生化剔除残差的值. 这些值都不大.

10.7.2 影响诊断

我们偶尔会发现, 数据的小子集对于拟合回归模型所起的作用有着与之不成比例的影响. 即参数估计或预测可能对有影响子集的依赖更甚于对主要数据的依赖. 我们想要找到这些强影响点并估计它们对模型的影响力. 如果这些强影响点起 "坏" 作用, 就应该将它们剔除. 另一方面, 这些点也可能没有错. 但如果这些点控制了主要模型特征, 我们还是想知道它们, 因为它可能影响到模型的使用. 本节将简述一些度量影响的有用方法.

1. 杠杆点

在 x 空间中点的分布对决定模型特征是重要的. 特别是, 远端的观测很可能对参数估计、预测值以及常规统计量有不成比例的杠杆作用.

帽子矩阵 $\boldsymbol{H}=\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ 在识别有较强影响的观测时是非常有用的. 如前面所述, 因为 $\mathrm{V}(\hat{\boldsymbol{y}})=\sigma^2\boldsymbol{H}$ 和 $\mathrm{V}(e)=\sigma^2(\boldsymbol{I}-\boldsymbol{H})$, 故 $\boldsymbol{H}$ 矩阵决定了 $\hat{\boldsymbol{y}}$ 和 $\boldsymbol{e}$ 的方差. $\boldsymbol{H}$ 的元素 h_{ij} 可被解释为 y_j 施加在 $\hat{y}_i$ 上杠杆作用的大小. 于是, 检查 $\boldsymbol{H}$ 的元素能够找出凭借它们在 x 空间中位置的优势而可能有的较强影响. 主要考虑对角元 h_{ii}. 因为 $\sum\limits_{i=1}^{n}h_{ii}=$ 秩 ($\boldsymbol{H}$) $=$ 秩 $(\boldsymbol{X})=p$, 故 $\boldsymbol{H}$ 矩阵对角元的平均大小是 p/n. 大体上说, 若一个对角元 h_{ii} 大于 $2\,p/n$, 则第 i 个观测就是高杠杆作用点. 对例 10.1 黏度模型而言, $2\,p/n=2(3)/16=0.375$. 表 10.3 给出了一阶模型的帽子矩阵对角元 h_{ii}, 因为 h_{ii} 中没有一个超过 0.375, 所以我们认为在这些数据中没有杠杆点.

2. 对回归系数的影响

可以用帽子矩阵对角元素识别出可能有强影响的点, 这种影响是由这些点在 x 空间所处的位置产生的. 在度量影响的大小时, 应该同时考虑点和响应变量的位置. Cook (1977, 1979) 建议, 将由所有 n 个点得到的最小二乘估计 $\hat{\boldsymbol{\beta}}$, 与剔除了第 i 点的数据得到的估计 $\hat{\boldsymbol{\beta}}_{(i)}$ 之间距离的平方, 作为第 i 个点影响的度量. 这个距离度量可写成

$$D_i=\frac{(\hat{\boldsymbol{\beta}}_{(i)}-\hat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}_{(i)}-\hat{\boldsymbol{\beta}})}{pMS_E}\quad i=1,2,\cdots,n \tag{10.54}$$

D_i 的一个合理界限值是一个单位. 即, 我们一般认为 $D_i>1$ 时有较大影响.

计算 D_i 统计量时实际所用的公式是

$$D_i=\frac{r_i^2}{p}\frac{\mathrm{V}[\hat{y}(x_i)]}{\mathrm{V}(e_i)}=\frac{r_i^2}{p}\frac{h_{ii}}{(1-h_{ii})}\quad i=1,2,\cdots,n \tag{10.55}$$

注意到, 除常数 p 以外, D_i 是第 i 个学生化残差的平方与 $h_{ii}\,/(1-h_{ii})$ 的乘积. 可以证明, 这个比值是从向量 $\boldsymbol{x}_i$ 到其余保留下来数据的中心点的距离. 于是, D_i 由两个部分构成, 一部分反映模型拟合第 i 个观测的程度, 另一部分度量了第 i 点与其余点的远离程度. 两部分都可能引起大的 D_i 值出现.

表 10.3 中列出了拟合例 10.1 黏度数据的回归模型的 D_i 值. 没有一个 D_i 值超过 1, 所以, 没有强烈证据表明这组数据中有强影响的观测值.

10.8 拟合不足检验

6.6 节说明了如何向一个 2^k 析因设计添加中心点, 以使实验者获得纯实验误差的估计. 这使得残差的平方和分解成两部分, 即

$$SS_E=SS_{\text{纯误差}}+SS_{\text{失拟}}$$

这里, $SS_{\text{纯误差}}$ 是纯误差平方和, $SS_{\text{失拟}}$ 是失拟平方和.

我们可以给出在回归模型中这个平方和分解式的推导. 假定在回归变量 $\boldsymbol{x}$ 的第 i 个水平 $\boldsymbol{x}_i$ 处有 n_i 个响应的观测, $i=1,2,\cdots,m$. 设 y_{ij} 表示在 $\boldsymbol{x}_i$ 处的响应的第 j 个观测, $i=1,2,\cdots,m$, $j=1,2,\cdots,n_i$. 共有 $\sum\limits_{i=1}^{m}n_i=n$ 个观测. 记第 (ij) 个残差为

$$y_{ij} - \hat{y}_i = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \hat{y}_i) \tag{10.56}$$

这里 $\bar{y}_i$ 是在 $\boldsymbol{x}_i$ 处 n_i 个观测的平均值. 对 (10.56) 式两边平方并对 i 与 j 求和, 得

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{m} n_i(\bar{y}_i - \hat{y}_i)^2 \tag{10.57}$$

(10.57) 式左边是通常所指的残差平方和, 右边两项是对纯误差和拟合不足的度量. 纯误差平方和

$$SS_{纯误差} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 \tag{10.58}$$

通过在 $\boldsymbol{x}$ 的每一水平上计算重复观测的修正平方和, 然后对 $\boldsymbol{x}$ 的 m 个水平求和获得. 若满足方差为常数的假定, 因为在每个 $\boldsymbol{x}_i$ 的水平上仅有 y 的变异性用于计算 $SS_{纯误差}$, 故这是纯误差的模型独立测量. 因为在每个水平 $\boldsymbol{x}_i$ 处, 纯误差的自由度为 n_i-1, 所以纯误差平方和的总自由度是

$$\sum_{i=1}^{m}(n_i - 1) = n - m \tag{10.59}$$

失拟平方和

$$SS_{失拟} = \sum_{i=1}^{m} n_i(\bar{y}_i - \hat{y}_i)^2 \tag{10.60}$$

是在每个 $\boldsymbol{x}_i$ 水平上的响应均值 $\bar{y}_i$ 与相应拟合值偏差平方的加权和. 如果拟合值 $\hat{y}_i$ 靠近相应的平均响应 $\bar{y}_i$, 就存在有力的证据表明回归函数是线性的. 如果 $\hat{y}_i$ 与 $\bar{y}_i$ 有很大偏离, 那么回归函数很可能不是线性的. 因为 $\boldsymbol{x}$ 有 m 个水平, 且模型在估计 p 个参数时损失了 p 个自由度, 故 $SS_{失拟}$ 有 $m-p$ 个自由度. 我们一般用从 SS_E 中减去 $SS_{纯误差}$ 的方法计算 $SS_{失拟}$.

检验统计量是

$$F_0 = \frac{SS_{失拟}/(m-p)}{SS_{纯误差}/(n-m)} = \frac{MS_{失拟}}{MS_{纯误差}} \tag{10.61}$$

$MS_{纯误差}$ 的期望是 σ^2, $MS_{失拟}$ 的期望是

$$\mathrm{E}(MS_{失拟}) = \sigma^2 + \frac{\sum_{i=1}^{m} n_i[E(y_i) - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij}]^2}{m-2} \tag{10.62}$$

如果真实的回归函数是线性的, 那么 $\mathrm{E}(y_i) = \beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}$, (10.62) 式中的第 2 项为零, 此时 $\mathrm{E}(MS_{失拟}) = \sigma^2$. 然而, 若回归函数不是线性的, 则 $\mathrm{E}(y_i) \neq \beta_0 + \sum_{j=1}^{k}\beta_j x_{ij}$, 于是 $\mathrm{E}(MS_{失拟}) > \sigma^2$. 因此, 如果真实的回归函数是线性的, 统计量 F_0 服从 $F_{m-p,n-m}$ 分布. 为了检验拟合是否不足, 先计算检验统计量 F_0 的值, 若 $F_0 > F_{\alpha,m-p,n-m}$, 则认为回归函数非线性.

这种检验方法可方便地合并到方差分析中. 如果回归函数是非线性, 则需放弃原来假定的模型, 并试图找出一个更合适的模型. 如果 $F_0 \leqslant F_{\alpha,m-p,n-m}$, 就不能证明拟合不足, 这时常常把 $MS_{纯误差}$ 与 $MS_{失拟}$ 合并起来估计 σ^2. 在中心点有重复试验的 2^2 分式析因设计的例 6.6 非常完整地说明了这个过程.

10.9 思考题

*10.1 纸张的抗拉强度与纸浆中硬木的含量有关. 在试验工厂中取 10 个样本, 所得数据如下表所示.

强度	硬木百分率	强度	硬木百分率
160	10	181	20
171	15	188	25
175	15	193	25
182	20	195	28
184	20	200	30

(a) 拟合强度关于硬木含量的一个线性回归模型.

(b) 检验 (a) 中所得模型的回归显著性.

(c) 求参数 β_1 的 95%置信区间.

10.2 工厂用蒸馏法从液态空气中提取氧、氮和氩. 可认为氧中的杂质百分率与空气中杂质含量有线性关系, 空气的杂质含量用 "污染读数" 来度量, 单位是百万分率 (ppm). 生产中的一个样本数据如下:

纯度 (%)	93.3	92.0	92.4	91.7	94.0	94.6	93.6	93.1	93.2	92.9	92.2	91.3	90.1	91.6	91.9
污染读数 (ppm)	1.10	1.45	1.36	1.59	1.08	0.75	1.20	0.99	0.83	1.22	1.47	1.81	2.03	1.75	1.68

(a) 对这些数据拟合一个线性回归模型.

(b) 检验回归显著性.

(c) 求参数 β_1 的 95%置信区间.

10.3 画出思考题 10.1 的残差图, 并论述模型的合适性.

10.4 画出思考题 10.2 的残差图, 并论述模型的合适性.

10.5 用思考题 10.1 的结果, 检验此回归模型的拟合不足.

10.6 研究轴承座的磨损问题, 轴承座的磨损 y 与 $x_1=$ 油的黏度及 $x_2=$ 负载有关. 已得下列数据:

y	x_1	x_2	y	x_1	x_2
193	1.6	851	91	43.0	1 201
230	15.5	816	113	33.0	1 357
172	22.0	1 058	125	40.0	1 115

(a) 对这些数据拟合一个多元线性回归模型.

(b) 检验回归显著性.

(c) 对每个模型参数计算 t 统计量. 你可以得到什么结论?

*10.7 用功率计测量的汽车发动机的制动马力被认为是一个以转/分为单位的发动机转速、燃料的行驶辛烷值以及发动机压缩的函数. 在实验室内进行一项实验, 收集到如下数据:

制动马力	转/分	行驶辛烷值	压缩	制动马力	转/分	行驶辛烷值	压缩
225	2 000	90	100	246	3 000	94	98
212	1 800	94	95	237	3 200	90	100
229	2 400	88	110	233	2 800	88	105
222	1 900	91	96	224	3 400	86	97
219	1 600	86	100	223	1 800	90	100
278	2 500	96	110	230	2 500	89	104

(a) 对这些数据拟合一个多元线性回归模型.

(b) 检验回归显著性. 你可以得到什么结论?

(c) 根据 t 检验的结果, 你在这个模型中需要所有的 3 个回归变量吗?

10.8 分析思考题 10.7 回归模型的残差并评价模型的合适性.

10.9 某一化学过程的产率和试剂的浓度与操作温度有关系. 进行一项实验, 结果如下.

产率	浓度	温度	产率	浓度	温度
81	1.00	150	79	1.00	150
89	1.00	180	87	1.00	180
83	2.00	150	84	2.00	150
91	2.00	180	90	2.00	180

(a) 假定我们想要对这些数据拟合一个主效应模型. 使用表中的数据求出 $\boldsymbol{X}'\boldsymbol{X}$ 矩阵.

(b) (a) 中所得到的矩阵是对角矩阵吗? 详述你的答案.

(c) 假定我们用 "通常的" 规范变量 $x_1 = \frac{\text{浓度}-1.5}{0.5}, x_2 = \frac{\text{温度}-165}{30}$ 写出模型. 用这两个规范变量的模型求出 $\boldsymbol{X}'\boldsymbol{X}$ 矩阵. 这个矩阵是对角矩阵吗? 详述你的答案.

(d) 定义新的规范变量 $x_1 = \frac{\text{浓度}-1.0}{1.0}, x_2 = \frac{\text{温度}-150}{30}$. 用这两个规范变量的模型求出 $\boldsymbol{X}'\boldsymbol{X}$ 矩阵. 这个矩阵是对角矩阵吗? 详述你的答案.

(e) 总结你在本题前几问中学到的有关规范变量的知识.

10.10 考虑例 6.2 中的 2^4 析因实验. 假定缺失了最后一个观测. 重新分析数据并作出结论. 这个结论与原来例中的结论相比有何不同?

10.11 考虑例 6.2 中的 2^4 析因实验. 假定缺失了最后两个观测. 重新分析数据并作出结论. 这个结论与原来例中的结论相比有何不同?

10.12 已知下列数据, 拟合二阶多项式回归模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \epsilon$$

y	x_1	x_2	y	x_1	x_2
26	1.0	1.0	62	1.5	2.0
24	1.0	1.0	100	0.5	3.0
175	1.5	4.0	26	1.0	1.5
160	1.5	4.0	30	0.5	1.5
163	1.5	4.0	70	1.0	2.5
55	0.5	2.0	71	0.5	2.5

拟合模型后, 检验回归的显著性.

10.13 (a) 考虑思考题 10.12 的二次回归模型. 对每个模型参数计算 t 统计量, 论述从这些量中得到的结论.

(b) 用附加平方和方法评估二次项 x_1^2, x_2^2 及 x_1x_2 对模型的价值.

10.14 **方差分析与回归的关系.** 任一方差分析模型可以用一般线性模型 $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ 来表示, 其中 $\boldsymbol{X}$ 矩阵由 0 和 1 组成. 证明单因子模型 $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i = 1, 2, 3$, $j = 1, 2, 3, 4$ 可写为一般线性模型的形式.

(a) 写出正规方程组 $(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$, 并与第 3 章对此模型求得的正规方程组相比较.

(b) 求 $\boldsymbol{X}'\boldsymbol{X}$ 的秩. 可以求得 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 吗?

(c) 设删去第一个正规方程并加上约束 $\sum\limits_{i=1}^{3} n\hat{\tau}_i = 0$. 所求得的方程组可解吗? 如果可解, 求出其解.

求回归平方和 $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}$, 并和单因子模型的处理平方和进行比较.

10.15 设拟合一直线并使 $\hat{\beta}_1$ 的方差尽可能小. 限取偶数个实验点, 这些点取在何处才能使 $V(\hat{\boldsymbol{\beta}}_1)$ 最小呢? (提示: 使用解此题的设计时要十分小心, 因为即使该设计可使 $V(\hat{\boldsymbol{\beta}}_1)$ 最小化, 也会有一些不良的性质. 例如, 可参阅 Myers and Montgomery(1995). 只有当你非常肯定真实的函数关系是线性关系时, 才可以考虑使用此设计.)

10.16 **加权最小二乘法**. 设拟合直线 $y = \beta_0 + \beta_1 x_1 + \varepsilon$, 但现在 y 的方差依赖于 x 的水平, 即,

$$V(y|x_i) = \sigma_i^2 = \frac{\sigma^2}{w_i} \quad i = 1, 2, \cdots, n$$

其中 w_i 是已知常数, 常称为权. 证明, 如果所选择的回归系数的估计可使加权误差平方和 $\sum_{i=1}^{n} w_i (y_i - \beta_0 - \beta_1 x_i)^2$ 最小化, 则所得的最小二乘正规方程是

$$\hat{\beta}_0 \sum_{i=1}^{n} w_i + \hat{\beta}_1 \sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{n} w_i y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} w_i x_i + \hat{\beta}_1 \sum_{i=1}^{n} w_i x_i^2 = \sum_{i=1}^{n} w_i x_i y_i$$

10.17 考虑例 10.5 中所讨论的 2_{IV}^{4-1} 设计.

(a) 假定你选择在这个例子的设计中补充一次试验. 求模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 + \beta_{34} x_3 x_4 + \varepsilon$$

的回归系数的方差和协方差 (忽略区组).

(b) 是否存在另一个分式设计中的任一试验可以分离别名 AB 与 CD?

(c) 假定你选择在例 10.5 的设计中补充 4 次试验. 求 (a) 中模型的回归系数的方差和协方差 (忽略区组).

(d) 考虑 (a) 和 (c), 你愿意用哪种补充试验的策略, 为什么?

10.18 考虑 2_{III}^{7-4} 设计. 假定在进行实验之后, 最大的观测效应是 $A + BD$, $B + AD$, $D + AB$. 你想在原设计中补充进行 4 次试验以分离这些效应的别名.

(a) 你愿意做哪 4 次试验?

(b) 求模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4 + \beta_{12} x_1 x_2 + \beta_{14} x_1 x_4 + \beta_{24} x_2 x_4 + \varepsilon$$

的回归系数的方差和协方差 (忽略区组).

(c) 用少于 4 次的附加试验是否有可能分离这些效应的别名?

第 11 章　响应曲面法与设计

本章纲要

11.1　响应曲面法引言
11.2　最速上升法
11.3　二阶响应曲面的分析
11.3.1　稳定点的位置
11.3.2　响应曲面的刻画
11.3.3　岭系统
11.3.4　多重响应
11.4　拟合响应曲面的实验设计
11.4.1　拟合一阶模型的设计
11.4.2　拟合二阶模型的设计
11.4.3　响应曲面设计的区组化
11.4.4　计算机生成 (最优) 设计
11.5　混料实验
11.6　调优运算
第 11 章补充材料
S11.1　最速上升法
S11.2　二阶响应曲面模型的正则形式
S11.3　中心复合设计的中心点
S11.4　面中心立方体的中心试验点
S11.5　旋转性方面的一点注解

11.1　响应曲面法引言

响应曲面法 (Response Surface Methodology, RSM) 是数学方法和统计方法结合的产物, 用于对感兴趣的响应受多个变量影响的问题进行建模和分析, 以优化这个响应. 例如, 设一位化学工程师想求出温度 (x_1) 和压强 (x_2) 的水平以使得过程的产率 (y) 达到最大值. 产率是温度水平和压强水平的函数, 比方说

$$y = f(x_1, x_2) + \varepsilon$$

其中 ε 表示响应 y 的观测误差或噪音. 如果记期望响应为 $\mathrm{E}(y) = f\ (x_1,\ x_2) = \eta$, 则由

$$\eta = f(x_1, x_2)$$

表示的曲面称为**响应曲面**.

通常像图 11.1 那样用图形来表示响应曲面, 其中 η 是对 x_1 水平和 x_2 水平画的. 前面已经看到过这样的响应曲面图形, 特别是在析因设计的各章中. 为了有助于目测响应曲面的形状, 经常像图 11.2 那样画出响应曲面的等高线. 在等高线图形中, 常数值的响应线画在 x_1, x_2 平面上. 每一条等高线对应于响应曲面的一个特定高度. 在前面章节中已看到了等高线图的作用.

在大多数 RSM 问题中, 响应和自变量之间的关系形式是未知的. 这样, RSM 的第一个步骤就是寻求 y 和自变量集合之间真实函数关系的一个合适的逼近式. 通常, 可用在自变量某一区域内的一个低阶多项式来逼近. 若响应适合用自变量的线性函数建模, 则近似函数是**一阶模型**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{11.1}$$

如果系统有弯曲, 则必须用更高阶的多项式, 例如**二阶模型**

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i<j}\sum \beta_{ij} x_i x_j + \varepsilon \tag{11.2}$$

几乎所有的 RSM 问题都用这两个模型中的一个或两个. 当然, 一个多项式模型不可能在自变量的整个空间上都是真实函数关系的合理近似式, 但在一个相对小的区域内通常做得很好.

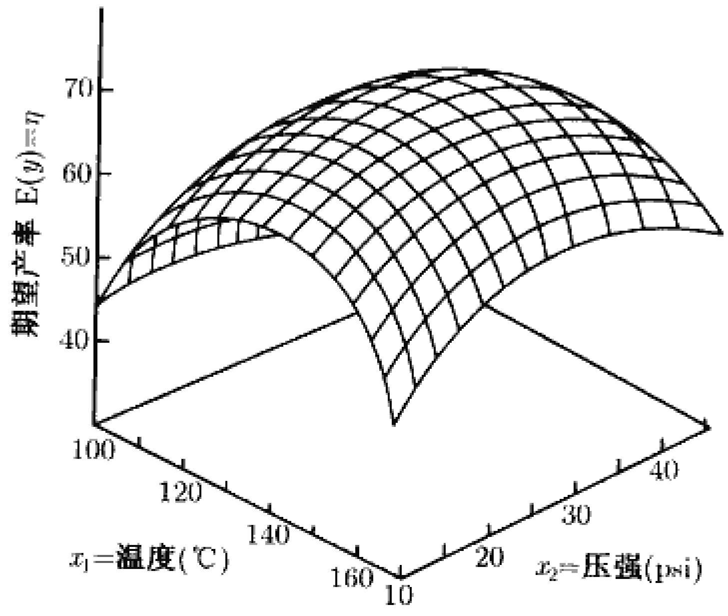


图 11.1　期望产率 (η) 作为温度 (x_1) 和压强 (x_2) 的函数的三维响应曲面

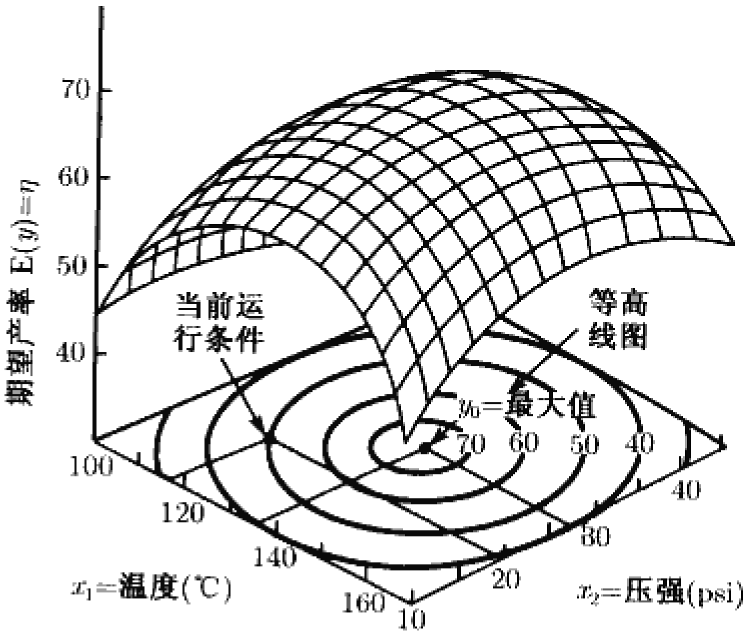


图 11.2　响应曲面的等高线图

第 10 章讨论的最小二乘方法可用来估计近似多项式的参数. 然后用拟合曲面进行响应曲面分析. 如果拟合曲面是真实响应函数的一个合适的近似式, 则拟合曲面的分析就近似地等价于实际系统的分析. 如果能恰当地利用实验设计来收集数据, 就能够最有效地估计模型参数. 拟合响应曲面的设计称为**响应曲面设计**. 11.4 节将讨论这些设计.

RSM 是一种**序贯方法**. 通常, 当我们在响应曲面上某个远离最优点的点时, 例如, 像图 11.3 中当前运行条件那样的点, 在此点处系统只有微小的弯曲, 从而用一阶模型是恰当的. 现在, 我们的目的是要引导实验者沿着改善系统的路径快速而有效地向**最优点**的附近区域前进. 一旦找到最优点的区域, 就可以用更精细的模型 (例如二阶模型) 进行分析以便确定最优点的位置. 由图 11.3 可见, 响应曲面的分析法可以想象为 "爬山" 一样, 山顶代表响应的最大值点. 如果真实的最优点是响应的最小值点, 则可设想为 "落进山谷".

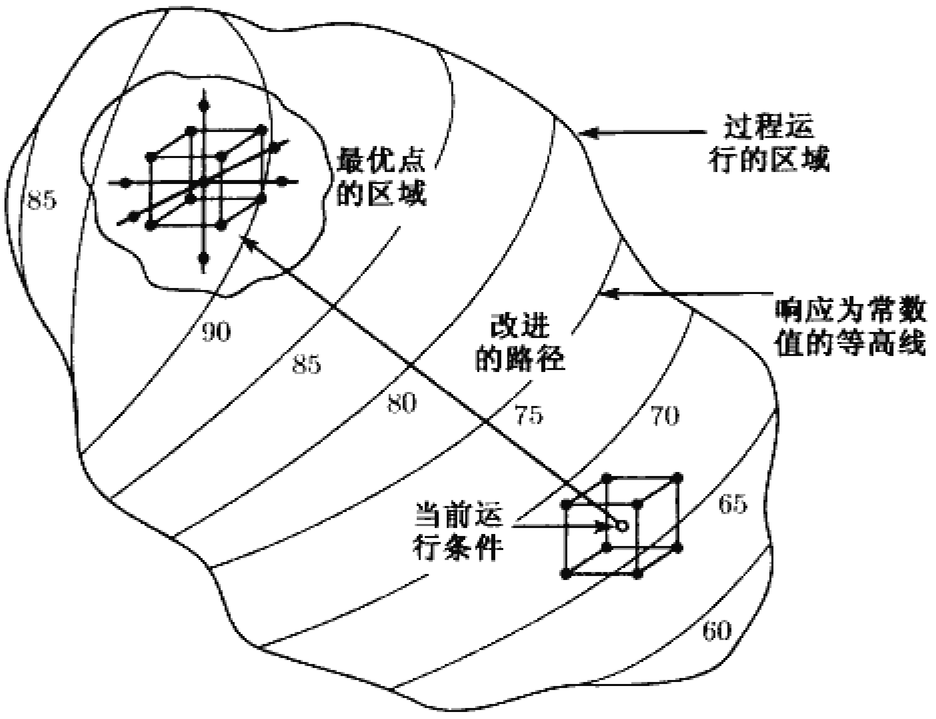


图 11.3 RSM 的序贯性质

RSM 的最终目的是确定系统的最优运行条件或确定因子空间中满足运行规范的区域. 有关 RSM 更全面的介绍见 Myers and Montgomery(2002), Khuri and Cornell(1996), 及 Box and Draper (1987). Myers 等人的综述论文也是有用的参考资料.

11.2 最速上升法

系统最优运行条件的初步估计常常远离实际的最优点. 在这种情况下, 实验者的目的是要快速地进入到最优点的附近区域. 我们希望利用既简单又经济有效的实验方法. 当远离最优点时, 通常假定在 x 的一个小区域范围内一阶模型是真实曲面的合适近似.

最速上升法是沿着响应有最大增量的方向逐步移动的方法. 当然, 如果求的是最小值, 则称为**最速下降法**. 所拟合的一阶模型是

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i \tag{11.3}$$

与一阶响应曲面相应的 $\hat{y}$ 的等高线, 是一组平行直线, 如图 11.4 所示. 最速上升的方向就是 $\hat{y}$ 增加得最快的方向. 这一方向平行于拟合响应曲面等高线的法线方向. 通常取通过感兴趣区域的中心并且垂直于拟合曲面等高线的直线为**最速上升路径**. 这样一来, 沿着路径的步长就和回归系数$\{\hat{\beta}_i\}$ 成正比. 实际的步长大小是由实验者根据工序知识或其他的实际考虑来确定的.

实验是沿着最速上升的路径进行的, 直到观测到的响应不再增加为止. 然后, 拟合一个新的一阶模型, 确定一条新的最速上升路径, 继续按上述方法进行. 最后, 实验者到达最优点的附近区域. 这通常由一阶模型的拟合不足来指出. 这时, 进行添加实验会求得最优点的更为精确的估计.

例 11.1 一位化学工程师要确定使过程产率最高的操作条件. 影响产率的两个可控变量是反应时间和反应温度. 工程师当前使用的操作条件是反应时间为 35 分钟, 温度为 155 ℉, 产率约为 40%. 因为此区域不大可能包含最优值, 于是她拟合一阶模型并应用最速上升法.

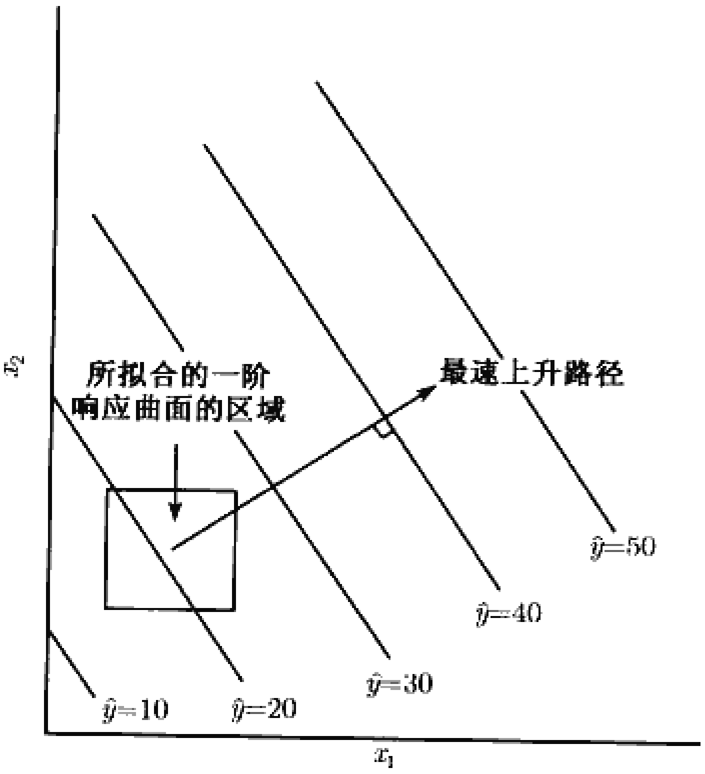


图 11.4 一阶响应曲面的等高线与最速上升路径

这位工程师认为，所拟合的一阶模型的探测区域应该是反应时间为 (30, 40) 分钟，反应温度为 (150, 160) ℉. 为简化计算，将自变量规范在 (-1, 1) 区间内. 于是，如果记 ξ_1 为**自然变量**时间，ξ_2 为**自然变量**温度，则**规范变量**是

$$x_1 = \frac{\xi_1 - 35}{5} \text{ 和 } x_2 = \frac{\xi_2 - 155}{5}$$

实验设计列在表 11.1 中. 用来收集这些数据的设计是增加 5 个中心点的 2^2 析因设计. 在中心点处的重复试验用于估计实验误差，并可以用于检测一阶模型的合适性. 而且，过程的当前运行条件也就在设计的中心点处.

表 11.1 拟合一阶模型的过程数据

自然变量		规范变量		响应
ξ_1	ξ_2	x_1	x_2	y
30	150	−1	−1	39.3
30	160	−1	1	40.0
40	150	1	−1	40.9
40	160	1	1	41.5
35	155	0	0	40.3
35	155	0	0	40.5
35	155	0	0	40.7
35	155	0	0	40.2
35	155	0	0	40.6

使用最小二乘法，以一阶模型来拟合这些数据. 用二水平设计的方法，可求得以规范变量表示的下列模型：

$$\hat{y} = 40.44 + 0.775x_1 + 0.325x_2$$

在沿着最速上升路径探测之前，应研究一阶模型的适合性. 有中心点的 2^2 设计使实验者能够

(1) 求出误差的一个估计量.

(2) 检测模型中的交互作用 (交叉乘积项).

(3) 检测二次效应 (弯曲性).

中心点处的重复试验观测值可用于计算误差的估计量:

$$\hat{\sigma}^2 = \frac{(40.3)^2 + (40.5)^2 + (40.7)^2 + (40.2)^2 + (40.6)^2 - (202.3)^2/5}{4} = 0.043\,0$$

一阶模型假定变量 x_1 和 x_2 对响应有可加效应. 变量间的交互作用可用已添加到模型中的交叉乘积项 x_1x_2 的系数 β_{12} 来表示. 此系数的最小二乘估计恰好是按普通 2^2 析因设计算得的交互作用效应的 1/2, 或

$$\hat{\beta}_{12} = \frac{1}{4}[(1 \times 39.3) + (1 \times 41.5) + (-1 \times 40.0) + (-1 \times 40.9)] = \frac{1}{4}(-0.1) = -0.025$$

单自由度的交互作用的平方和是

$$SS_{\text{交互作用}} = \frac{(-0.1)^2}{4} = 0.002\,5$$

比较 $SS_{\text{交互作用}}$ 和 $\hat{\sigma}^2$, 得到下列拟合不足统计量:

$$F = \frac{SS_{\text{交互作用}}}{\hat{\sigma}^2} = \frac{0.002\,5}{0.043\,0} = 0.058$$

它很小, 表示交互作用可以忽略.

对直线模型适合性的另一种检测方法, 是应用 6.6 节中所介绍的对纯二次弯曲效应的检测. 回忆一下它的构成, 也就是, 比较设计的析因部分 4 个点处的平均响应, 即 $\bar{y}_F = 40.425$, 以及在设计的中心点处的平均响应 $\bar{y}_C = 40.46$. 如果在真实响应函数中存在二次弯曲性, 则 $\bar{y}_F - \bar{y}_C$ 是这种弯曲性的度量. 如果 β_{11} 与 β_{22} 是 "纯二次" 项 x_1^2 与 x_2^2 的系数, 则 $\bar{y}_F - \bar{y}_C$ 是 $\beta_{11} + \beta_{22}$ 的一个估计. 在我们的例子中, 纯二次项的一个估计是

$$\hat{\beta}_{11} + \hat{\beta}_{22} = \bar{y}_F - \bar{y}_C = 40.425 - 40.46 = -0.035$$

与零假设 H_0: $\beta_{11} + \beta_{22} = 0$ 有关的单自由度的平方和是

$$SS_{\text{纯二次}} = \frac{n_F n_C(\bar{y}_F - \bar{y}_C)^2}{n_F + n_C} = \frac{(4)(5)(-0.035)^2}{4+5} = 0.002\,7$$

其中 n_F 与 n_C 分别是析因部分的点数和中心点数. 因为

$$F = \frac{SS_{\text{纯二次}}}{\hat{\sigma}^2} = \frac{0.002\,7}{0.043\,0} = 0.063$$

很小, 没有显示出纯二次项的影响.

此模型的方差分析概括在表 11.2 中. 交互作用和弯曲性的检验都不显著, 而总回归的 F 检验是显著的. 此外, $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的标准差是

$$se(\hat{\beta}_i) = \sqrt{\frac{MS_E}{4}} = \sqrt{\frac{\hat{\sigma}^2}{4}} = \sqrt{\frac{0.043\,0}{4}} = 0.10 \quad i = 1, 2$$

回归系数 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 相对于它们的标准差都较大. 此时我们没有理由怀疑一阶模型的合适性.

表 11.2 一阶模型的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
回归 (β_1, β_2)	2.825 0	2	1.412 5	47.83	0.000 2
残差	0.177 2	6			
(交互作用)	(0.002 5)	1	0.002 5	0.058	0.821 5
(纯二次)	(0.002 7)	1	0.002 7	0.063	0.814 2
(纯误差)	(0.172 0)	4	0.043 0		
总和	3.002 2	8			

要离开设计中心 —— 点 ($x_1 = 0, x_2 = 0$)—— 沿最速上升路径移动, 对应于沿 x_2 方向每移动 0.325 个单位, 则应沿 x_1 方向移动 0.775 个单位. 于是, 最速上升路径经过点 ($x_1 = 0, x_2 = 0$) 且斜率为 0.325/0.775. 工程师决定用 5 分钟反应时间作为基本步长. 由 ξ_1 与 x_1 之间的关系式, 知道 5 分钟反应时间等价于规范变量 x_1 的步长为 $\Delta x_1 = 1$. 因此, 沿最速上升路径的步长是 $\Delta x_1 = 1.000\ 0$ 和 $\Delta x_2 = (0.325/0.775)\Delta x_1 = 0.42$.

工程师计算了沿此路径的点, 并观测了在这些点处的产率直至响应有下降为止. 其结果见表 11.3, 表中既列出了规范变量, 也列出了自然变量. 虽然规范变量在数学上容易计算, 但在过程运行中必须用自然变量. 图 11.5 画出了沿最速上升路径的每一步处的产率图. 一直到第 10 步所观测到的响应都是增加的; 但是, 这以后的每一步收率都是减少的, 因此, 另一个一阶模型应该在点 ($\xi_1 = 85, \xi_2 = 175$) 的附近区域进行拟合.

表 11.3　例 11.1 的最速上升实验

步长	规范变量		自然变量		响应 y
	x_1	x_2	ξ_1	ξ_2	
原点	0	0	35	155	
Δ	1.00	0.42	5	2	
原点 + Δ	1.00	0.42	40	157	41.0
原点 + 2Δ	2.00	0.84	45	159	42.9
原点 + 3Δ	3.00	1.26	50	161	47.1
原点 + 4Δ	4.00	1.68	55	163	49.7
原点 + 5Δ	5.00	2.10	60	165	53.8
原点 + 6Δ	6.00	2.52	65	167	59.9
原点 + 7Δ	7.00	2.94	70	169	65.0
原点 + 8Δ	8.00	3.36	75	171	70.4
原点 + 9Δ	9.00	3.78	80	173	77.6
原点 + 10Δ	10.00	4.20	85	175	80.3
原点 + 11Δ	11.00	4.62	90	177	76.2
原点 + 12Δ	12.00	5.04	95	179	75.1

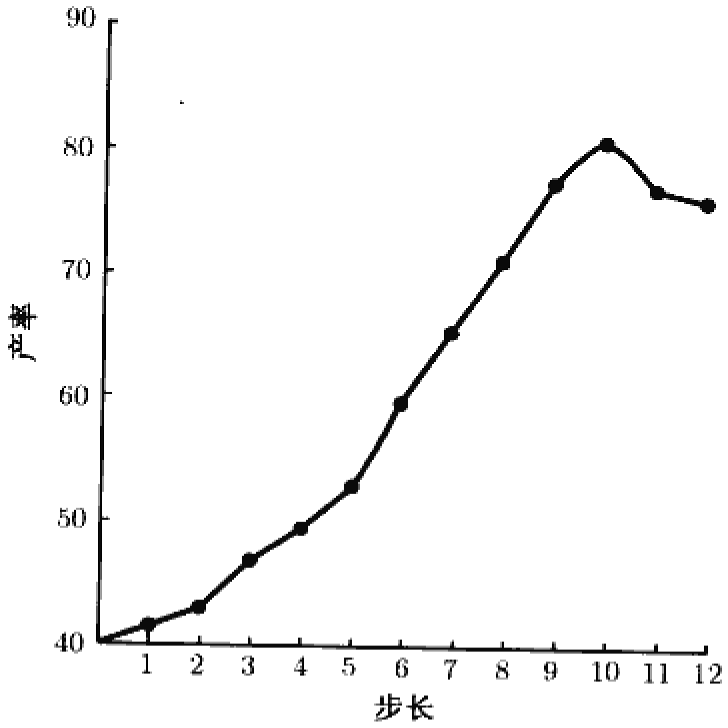


图 11.5　例 11.1 中沿最速上升路径的产率与步长的关系图

一个新的一阶模型在点 (ξ_1=85, ξ_2=175) 附近拟合. 探测的区域对 ξ_1 是 [80, 90], 对 ξ_2 是 [170, 180],

于是, 规范变量是

$$x_1=\frac{\xi_1-85}{5} \text{ 和 } x_2=\frac{\xi_2-175}{5}$$

再次用有 5 个中心点的 2^2 设计. 实验设计列在表 11.4 中.

表 11.4 第 2 个一阶模型的数据

自然变量		规范变量		响应 y
ξ_1	ξ_2	x_1	x_2	
80	170	−1	−1	76.5
80	180	−1	1	77.0
90	170	1	−1	78.0
90	180	1	1	79.5
85	175	0	0	79.9
85	175	0	0	80.3
85	175	0	0	80.0
85	175	0	0	79.7
85	175	0	0	79.8

拟合表 11.4 的规范数据的一阶模型是

$$\hat{y}=78.97+1.00x_1+0.50x_2$$

此模型的方差分析, 包括交互作用和纯二次项的检测, 如表 11.5 所示. 交互作用和纯二次项的检测表明, 一阶模型不是合适的近似. 真实曲面的弯曲性指明了我们已接近最优点. 为更精细地确定最优点, 在该点必须做进一步的分析.

表 11.5 第 2 个一阶模型的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
回归 (β_1, β_2)	5.00	2			
残差	11.120 0	6			
(交互作用)	(0.250 0)	1	0.250 0	4.72	0.095 5
(纯二次)	(10.658 0)	1	10.658 0	201.09	0.000 1
(纯误差)	(0.212 0)	4	0.053 0		
总和	16.120 0	8			

由例 11.1 可见, *最速上升路径与拟合的一阶模型*

$$\hat{y}=\hat{\beta}_0+\sum_{i=1}^{k}\hat{\beta}_i x_i$$

的回归系数的符号和大小成比例. 容易给出一个一般算法, 以确定最速上升路径上点的坐标. 假定 $x_1=x_2=\cdots=x_k=0$ 是基点或原点. 则

(1) 选取一个过程变量的步长, 比方说 Δx_j. 通常, 选取我们最了解的变量, 或选取其回归系数的绝对值 $|\hat{\beta}_j|$ 最大的变量.

(2) 其他变量的步长是

$$\Delta x_i=\frac{\hat{\beta}_i}{\hat{\beta}_j/\Delta x_j},\qquad i=1,2,\cdots,k;\quad i\neq j$$

(3) 将规范变量的 Δx_i 转换至自然变量.

为了说明起见, 考虑例 11.1 最速上升路径的计算. 因为 x_1 有最大的回归系数, 选取反应时间作为上述方法的步骤 1 中的变量. 5 分钟反应时间是步长 (根据工序知识). 用规范变量的说法, 也就是 $\Delta x_1 = 1.0$, 因此, 由步骤 2, 温度的步长是

$$\Delta x_2 = \frac{\hat{\beta}_2}{\hat{\beta}_1/\Delta x_1} = \frac{0.325}{(0.775/1.0)} = 0.42$$

为了将规范步长 ($\Delta x_1 = 1.0$, $\Delta x_2 = 0.42$) 转换为时间和温度的自然单位, 用关系式

$$\Delta x_1 = \frac{\Delta \xi_1}{5} \text{ 和 } \Delta x_2 = \frac{\Delta \xi_2}{5}$$

其结果为

$$\Delta \xi_1 = \Delta x_1(5) = 1.0(5) = 5 \text{ 分}$$

$$\Delta \xi_2 = \Delta x_2(5) = 0.42(5) = 2\ ^\circ\text{F}$$

11.3　二阶响应曲面的分析

当实验者相对接近最优点时, 通常需要一个具有弯曲性的模型来逼近响应. 在大多数情况下, 二阶模型

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i<j}\sum \beta_{ij} x_i x_j + \varepsilon \tag{11.4}$$

是合适的. 本节将指出怎样利用这一拟合模型来寻求 x 运行条件的最优集合并指出响应曲面的特征性质.

11.3.1　稳定点的位置

假设我们想要求出 x_1, x_2, $\cdots$, x_k 的水平, 使之能够最优化所预测的响应. 这个点 (x_1, x_2, $\cdots$, x_k) 如果存在的话, 它应使偏导数 $\partial \hat{y}/\partial x_1 = \partial \hat{y}/\partial x_2 = \cdots = \partial \hat{y}/\partial x_k = 0$. 记它为 $(x_{1,s}, x_{2,s}, \cdots, x_{k,s})$, 称为**稳定点**(stationary point). 稳定点可以是: (1) **响应的最大值**点, (2) **响应的最小值**点, (3) **鞍点**(saddle point). 这 3 种可能性见图 11.6、图 11.7 和图 11.8.

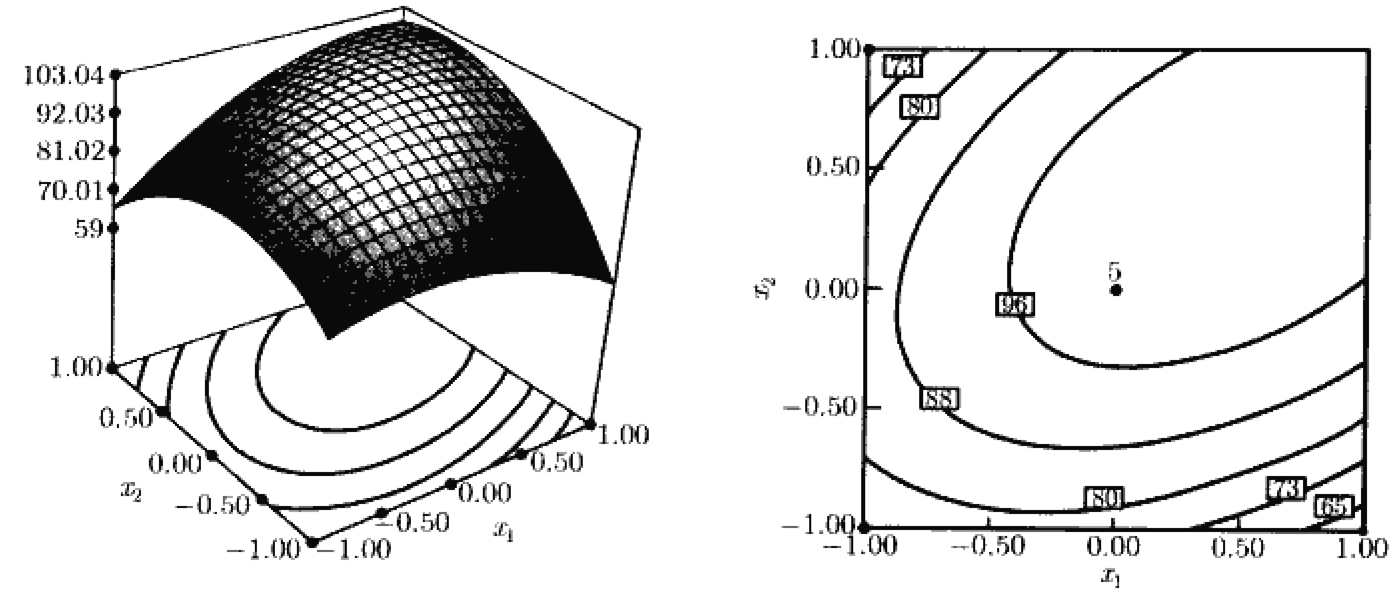
等高线图在响应曲面的研究中起着非常重要的作用. 根据用于响应曲面分析的计算机软件产生的等高线图, 实验者可以描述曲面形状的特征并以较好的精度找到最优点.

可以求出稳定点位置的一般数学解. 将所拟合的二阶模型写成矩阵记号形式, 有

$$\hat{y} = \hat{\beta}_0 + \boldsymbol{x}'\boldsymbol{b} + \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} \tag{11.5}$$

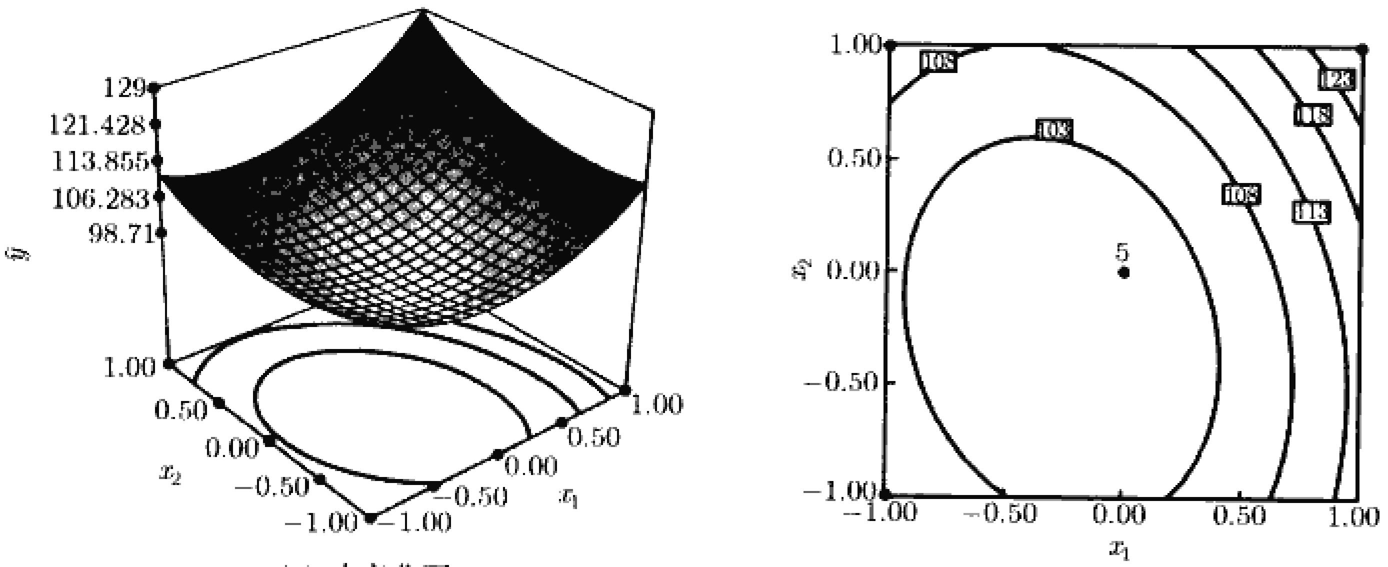
其中

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \quad \boldsymbol{B} = \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/2 & \cdots & \hat{\beta}_{1k}/2 \\ & \hat{\beta}_{22} & \cdots & \hat{\beta}_{2k}/2 \\ & & \ddots & \\ \text{对称} & & & \hat{\beta}_{kk} \end{bmatrix}$$



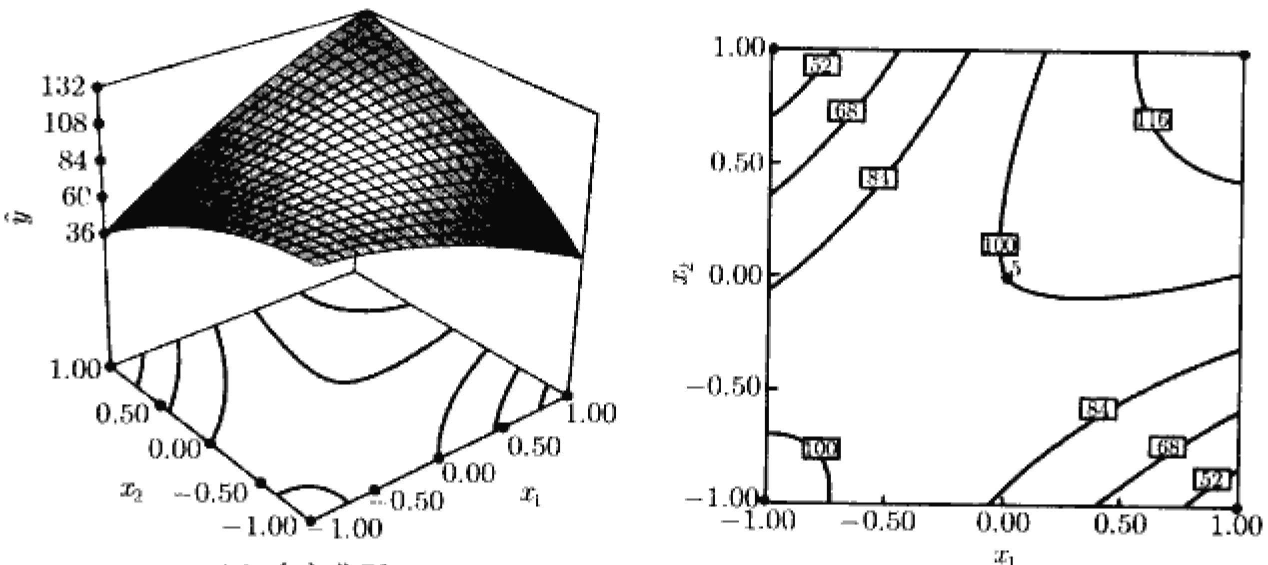
(a) 响应曲面　　(b) 等高线图

图 11.6　具有最大值的曲面的响应曲面和等高线图



(a) 响应曲面　　(b) 等高线图

图 11.7　具有最小值的曲面的响应曲面和等高线图



(a) 响应曲面　　(b) 等高线图

图 11.8　有鞍点 (或极小极大) 的曲面的响应曲面和等高线图

也就是说, $\boldsymbol{b}$ 是一阶回归系数的一个 $(k\times1)$ 向量, $\boldsymbol{B}$ 是 $(k\times k)$ 对称矩阵, 其主对角线元素是纯二次系数 $(\hat{\beta}_{ii})$, 非对角元素是混合二次系数 $(\hat{\beta}_{ij}, i\neq j)$ 的 1/2. $\hat{y}$ 关于向量 $\boldsymbol{x}$ 的元素的导数等

于 $\boldsymbol{0}$ 就是

$$\frac{\partial \hat{y}}{\partial \boldsymbol{x}} = \boldsymbol{b} + 2\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0} \tag{11.6}$$

稳定点是 (11.6) 式的解, 即

$$\boldsymbol{x}_s = -\frac{1}{2}\boldsymbol{B}^{-1}\boldsymbol{b} \tag{11.7}$$

然后将 (11.7) 式代入 (11.5) 式, 求得在稳定点处的预测响应为

$$\hat{y}_s = \hat{\beta}_0 + \frac{1}{2}\boldsymbol{x}_s'\boldsymbol{b} \tag{11.8}$$

11.3.2 响应曲面的刻画

一旦求出稳定点, 通常必需刻画出响应曲面在这一点最邻近区域内的特征. 所谓**刻画**, 意思是要确定该稳定点究竟是响应的最大值点、最小值点还是鞍点. 通常我们也想研究响应对变量 $x_1, x_2, \cdots, x_k$ 的相对敏感度.

正如前面已提及的, 要做到这一点, 最直接的方法是去考察所拟合模型的**等高线图**. 如果只有 2 个或 3 个过程变量 (x), 则等高线图的构造和解释相对容易. 但是, 即使有相对少量的变量, 也可以用称为**正则分析**的更为正规的分析方法.

首先利用坐标变换将模型放入一个新的坐标系, 它的原点在稳定点 $\boldsymbol{x}_s$ 处, 然后旋转坐标系, 直至它们与所拟合响应曲面的主轴平行为止, 此变换如图 11.9 所示. 可以证明, 这样得出的拟合模型是

$$\hat{y} = \hat{y}_s + \lambda_1 w_1^2 + \lambda_2 w_2^2 + \cdots + \lambda_k w_k^2 \tag{11.9}$$

其中$\{w_i\}$是变换后的自变量, $\{\lambda_i\}$ 是常数. (11.9) 式称为模型的**正则形式**. 而且$\{\lambda_i\}$ 恰好是矩阵 $\boldsymbol{B}$ 的**特征值或特征根**.

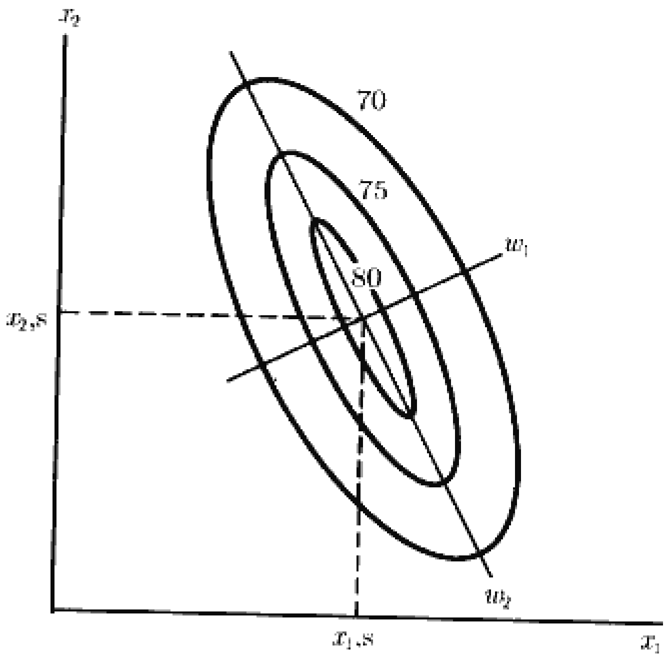


图 11.9 二阶模型的正则形式

响应曲面的性质可以由稳定点与$\{\lambda_i\}$ 的符号和大小来确定. 首先, 设稳定点在拟合二阶模型所探测的区域之内, 如果$\{\lambda_i\}$ 都是正的, 则 $\boldsymbol{x}_s$ 是响应的最小值点; 如果$\{\lambda_i\}$ 都是负的, 则 $\boldsymbol{x}_s$ 是响应的最大值点; 如果$\{\lambda_i\}$ 有不同的符号, 则 $\boldsymbol{x}_s$ 是鞍点. 此外, 当 $|\lambda_i|$ 最大时, w_i 的方向

是曲面最陡的方向. 例如, 图 11.9 画出了一个系统, 其 $\boldsymbol{x}_s$ 是最大值点 (λ_1 和 λ_2 都是负的), 且 $|\lambda_1| > |\lambda_2|$.

例 11.2 我们继续分析例 11.1 的化学过程. 仅用表 11.4 的设计不能拟合变量 x_1 和 x_2 的二阶模型. 实验者决定以足够多的点增大这个设计使得能拟合一个二阶模型[①]. 她在 ($x_1 = 0$, $x_2 = \pm1.414$) 和 ($x_1 = \pm1.414$, $x_2 = 0$) 处得到 4 个观测值. 完整的实验列在表 11.6 中, 设计显示在图 11.10 中. 此设计称为**中心复合设计** (CCD) 并将在 11.4.2 节中进行更为详细的讨论. 在研究的第二阶段, 考虑另两个响应: 产品的黏度和分子量. 这两个响应也列在表 11.6 中.

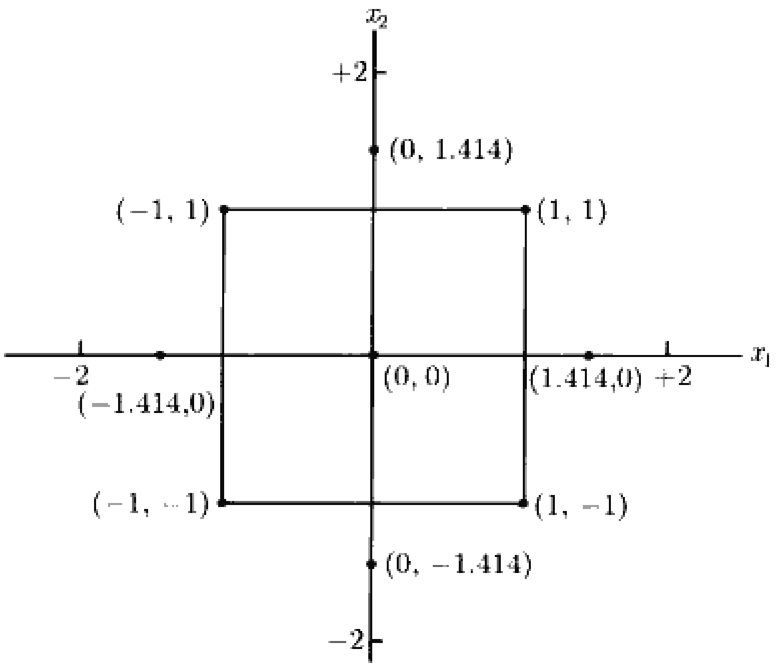


图 11.10 例 11.2 的中心复合设计

表 11.6 例 11.2 的中心复合设计

自然变量		规范变量		响应		
ξ_1	ξ_2	x_1	x_2	y_1 (产率)	y_2 (黏度)	y_3(分子量)
80	170	−1	−1	76.5	62	2 940
80	180	−1	1	77.0	60	3 470
90	170	1	−1	78.0	66	3 680
90	180	1	1	79.5	59	3 890
85	175	0	0	79.9	72	3 480
85	175	0	0	80.3	69	3 200
85	175	0	0	80.0	68	3 410
85	175	0	0	79.7	70	3 290
85	175	0	0	79.8	71	3 500
92.07	175	1.414	0	78.4	68	3 360
77.93	175	−1.414	0	75.6	71	3 020
85	182.07	0	1.414	78.5	58	3 630
85	167.93	0	−1.414	77.0	57	3 150

这里主要考虑拟合产率响应 y_1 的二次模型 (其他响应在 11.3.4 节中讨论). 通常使用计算机软件来拟合响应曲面并画出等高线图. 表 11.7 包含了 Design-Expert 的输出结果. 在表中可以看到, 软件包首先计

① 工程师在做原来的 9 个观测值的试验的同时, 做附加的 4 个观测值的试验. 如果两组试验间的时间间隔太长, 则必须划分区组. 响应曲面设计的区组化在 11.4.3 节中讨论.

算了模型中的线性、二次、三次项的 "序贯或额外平方和" (在 3 次模型中有一个关于别名的警告信息, 因为 CCD 中并未包含足以支持完全 3 次模型的试验次数). 由于二次项有小的 P 值, 我们决定采用二阶模型来拟合产率响应. 计算机输出的最终模型既有规范变量形式的模型, 也有自然 (或实际) 因子水平的模型.

表 11.7 例 11.2 对产率响应的拟合模型的 Design-Expert 计算机输出

Response: yield

WARNING: The Cubic Model is Aliased!

Sequential Model Sum of Squares

Source	Sum of Squares	DF	Mean Square	F Value	Prob>F	
Mean	80062.16	1	80062.16			
Linear	10.04	2	5.02	2.69	0.1166	
2FI	0.25	1	0.25	0.12	0.7350	
Quadratic	*17.95*	*2*	*8.98*	*126.88*	*<0.001*	*Suggested*
Cubic	2.042E-003	2	1.021E-003	0.010	0.9897	Aliased
Residual	0.49	5	0.099			
Total	80090.90	13	6160.84			

''Sequential Model Sum of Squares'': Select the highest order polynomial where the additional terms are significant.

Lack of Fit Tests

Source	Sum of Squares	DF	Mean Square	F Value	Prob>F	
Linear	18.49	6	3.08	58.14	0.0008	
2FI	18.24	5	3.65	68.82	0.0006	
Quadratic	*0.28*	*3*	*0.094*	*1.78*	*0.2897*	*Suggested*
Cubic	0.28	1	0.28	5.31	0.0826	Aliased
Pure Error	0.21	4	0.053			

''Lack of Fit Tests'': Want the selected model to have insignificant lack-of-fit.

Model Summary Statistics

Source	Std. Dev.	R-Squared	Adjusted R-Squared	Predicted R-Squared	PRESS	
Linear	1.37	0.3494	0.2193	−0.0435	29.99	
2FI	1.43	0.3581	0.1441	−0.2730	36.59	
Quadratic	*0.27*	*0.9828*	*0.9705*	*0.9184*	*2.35*	*Suggested*
Cubic	0.31	0.9828	0.9588	0.3622	18.33	Aliased

''Model Summary Statistics'': Focus on the model minimizing the ''PRESS'', or equivalently maximizing the ''PRED R-SQR''.

Response: yield

ANOVA for Response Surface Quadratic Model

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob>F

(续)

Model	28.25	5	5.65	79.85	<0.0001
A	*7.92*	*1*	*7.92*	*111.93*	*<0.0001*
B	*2.12*	*1*	*2.12*	*30.01*	*0.0009*
A^2	*13.18*	*1*	*13.18*	*186.22*	*<0.0001*
B^2	*6.97*	*1*	*6.97*	*98.56*	*<0.0001*
AB	*0.25*	*1*	*0.25*	*3.53*	*0.1022*
Residual	0.50	7	0.071		
Lack of Fit	*0.28*	*3*	*0.094*	*1.78*	*0.2897*
Pure Error	*0.21*	*4*	*0.053*		
Cor Total	28.74	12			

Std. Dev.	0.27	R-Squared	0.9828
Mean	78.48	Adj R-Squared	0.9705
C.V.	0.34	Pred R-Squared	0.9184
PRESS	2.35	Adeq Precision	23.018

Factor	Coefficient Estimate	DF	Standard Error	95% CI Low	95% CI High	VIF
Intercept	79.94	1	0.12	79.66	80.22	
A-time	0.99	1	0.094	0.77	1.22	1.00
B-temp	0.52	1	0.094	0.29	0.74	1.00
A^2	-1.38	1	0.10	-1.61	-1.14	1.02
B^2	-1.00	1	0.10	-1.24	-0.76	1.02
AB	0.25	1	0.13	-0.064	0.56	1.00

Final Equation in Terms of Coded Factors:

$$\text{yield} = +79.94 + 0.99 * A + 0.52 * B - 1.38 * A^2 - 1.00 * B^2 + 0.25 * A * B$$

Final Equation in Terms of Actual Factors:

$$\text{yield} = -1430.52285 + 7.80749 * \text{time} + 13.27053 * \text{temp} - 0.055050 * \text{time}^2 - 0.040050 * \text{temp}^2 + 0.010000 * \text{time} * \text{temp}$$

Diagnostics Case Statistics

(续)

Run Order	Standard Order	Actual Value	Predicted Value	Residual	Leverage	Student Residual	Cook's Distance	Outlier t
8	1	76.50	76.30	0.20	0.625	1.213	0.409	1.264
6	2	78.00	77.79	0.21	0.625	1.275	0.452	1.347
9	3	77.00	76.83	0.17	0.625	1.027	0.293	1.032
11	4	79.50	79.32	0.18	0.625	1.089	0.329	1.106
12	5	75.60	75.78	−0.18	0.625	−1.107	0.341	−1.129
10	6	78.40	78.59	−0.19	0.625	−1.195	0.396	−1.240
7	7	77.00	77.21	−0.21	0.625	−1.283	0.457	−1.358
1	8	78.50	78.67	−0.17	0.625	−1.019	0.289	−1.023
5	9	79.90	79.94	−0.040	0.200	−0.168	0.001	−0.156
3	10	80.30	79.94	0.36	0.200	1.513	0.095	1.708
13	11	80.00	79.94	0.060	0.200	0.252	0.003	0.235
2	12	79.70	79.94	−0.24	0.200	−1.009	0.042	−1.010
4	13	79.80	79.94	−0.14	0.200	−0.588	0.014	−0.559

图 11.11 显示了由过程变量时间和温度表示的产率响应的三维响应曲面图形及对应的等高线图. 考察这两张图形, 可以相对容易地看出, 最优点非常接近 175 ℉和 85 分钟反应时间, 而且响应在这一点达到最大值. 从等高线图可以看出, 过程对反应时间的变化比对温度的变化更为敏感一些.

用 (11.7) 式中的通解也能求出稳定点的位置. 注意到

$$\boldsymbol{b}=\begin{bmatrix}0.995\\0.515\end{bmatrix}\quad \boldsymbol{B}=\begin{bmatrix}-1.376 & 0.125\,0\\0.125\,0 & -1.001\end{bmatrix}$$

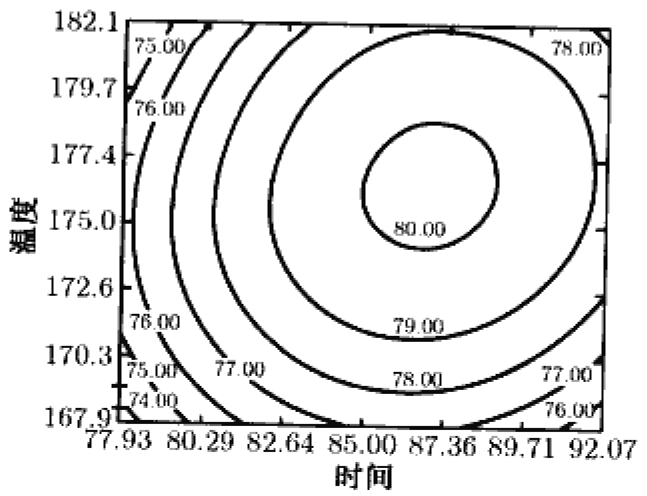
由 (11.7) 式, 稳定点是

$$\boldsymbol{x}_s=-\frac{1}{2}\boldsymbol{B}^{-1}\boldsymbol{b}=-\frac{1}{2}\begin{bmatrix}-0.734\,5 & -0.091\,7\\-0.091\,7 & -1.009\,6\end{bmatrix}\begin{bmatrix}0.995\\0.515\end{bmatrix}=\begin{bmatrix}0.389\\0.306\end{bmatrix}$$

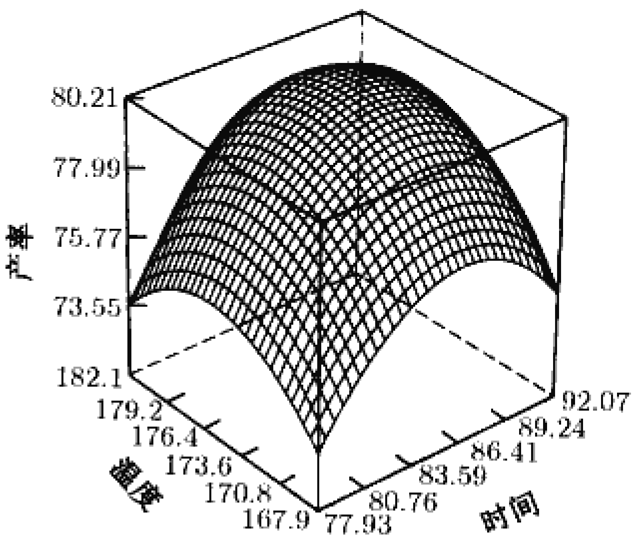
即 $x_{1,s}=0.389$, $x_{2,s}=0.306$. 转换为自然变量, 稳定点满足

$$0.389=\frac{\xi_1-85}{5}\qquad 0.306=\frac{\xi_2-175}{5}$$

得 $\xi_1=86.95\approx 87$ 分钟反应时间, $\xi_2=176.53\approx 176.5$ ℉. 这非常靠近在图 11.11 的等高线图中目测到的稳定点. 使用 (11.8) 式, 求得在稳定点处的预测响应是 $\hat{y}_s=80.21$.



(a) 等高线图



(b) 响应曲面图

图 11.11 例 11.2 产率响应的等高线图和响应曲面图

也可以用本节描述的正则分析法来刻画响应曲面. 首先, 需要将所拟合的模型用正则形式 [(11.9) 式] 表示出来. 特征值 λ_1 和 λ_2 是行列式方程

$$|\boldsymbol{B}-\lambda\boldsymbol{I}|=0$$

$$\begin{vmatrix} -1.376-\lambda & 0.125\ 0 \\ 0.125\ 0 & -1.001-\lambda \end{vmatrix}=0$$

的根, 把上面的行列式方程化简即得

$$\lambda^2+2.378\ 8\lambda+1.363\ 9=0$$

此二次方程的根是 $\lambda_1=-0.964\ 1$ 和 $\lambda_2=-1.414\ 7$. 于是, 所拟合的模型的正则形式是

$$\hat{y}=80.21-0.964\ 1w_1^2-1.414\ 7w_2^2$$

因为 λ_1 和 λ_2 都是负的, 且稳定点在探测区域内. 我们的结论是, 稳定点是最大值点.

在有些 RSM 问题中, 必须求出**正则变量** $\{w_i\}$ 和**设计变量** $\{x_i\}$ 之间的关系. 当过程不能在稳定点处运行时, 尤其需要这样做. 作为说明, 设在例 11.2 中, 过程不能在 ξ_1=87 分钟和 ξ_2=176.5 ℉ 处运行, 因为这一因子组合导致成本过大. 现在想从稳定点 "返回" 至一个较低成本的点又不至于在产率上有较大的损失. 模型的正则形式显示出曲面沿 w_1 方向的产率损失较小. 正则形式的研究需要将 (w_1, w_2) 空间中的点变换为 (x_1, x_2) 空间中的点.

一般说来, 变量 $\boldsymbol{x}$ 与正则变量 $\boldsymbol{w}$ 之间的关系是

$$\boldsymbol{w}=\boldsymbol{M}'(\boldsymbol{x}-\boldsymbol{x}_s)$$

其中 $\boldsymbol{M}$ 是 $(k\times k)$ 正交矩阵. $\boldsymbol{M}$ 的列是与$\{\lambda_i\}$ 对应的标准化特征向量. 也就是说, 如果 $\boldsymbol{m}_i$ 是 $\boldsymbol{M}$ 的第 i 列, 则 $\boldsymbol{m}_i$ 是

$$(\boldsymbol{B}-\lambda_i\boldsymbol{I})\boldsymbol{m}_i=\boldsymbol{0} \tag{11.10}$$

的解, 而且 $\sum_{j=1}^{k} m_{ji}^2=1$.

我们用例 11.2 的二阶拟合模型来说明这一方法. 对 $\lambda_1=-0.964\ 1$, (11.10) 式成为

$$\begin{bmatrix} -1.376+0.964\ 1 & 0.125\ 0 \\ 0.125\ 0 & -1.001+0.964\ 1 \end{bmatrix}\begin{bmatrix} m_{11} \\ m_{21} \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

或

$$-0.412\ 9m_{11}+0.125\ 0m_{21}=0$$

$$0.125\ 0m_{11}-0.037\ 7m_{21}=0$$

我们要求此方程组的标准化解, 即 $m_{11}^2+m_{21}^2=1$ 的解. 此方程组没有唯一的解, 所以, 最方便的方法是对一个未知数指定一个任意值, 解此方程组, 然后将解标准化. 令 $m_{21}^*=1$, 求得 $m_{11}^*=0.302\ 7$. 要使这一解标准化, 将 m_{11}^* 和 m_{21}^* 除以

$$\sqrt{(m_{11}^*)^2+(m_{21}^*)^2}=\sqrt{0.302\ 7^2+1^2}=1.044\ 8$$

这就得出标准化解:

$$m_{11}=\frac{m_{11}^*}{1.044\ 8}=\frac{0.302\ 7}{1.044\ 8}=0.289\ 7$$

$$m_{21}=\frac{m_{21}^*}{1.044\ 8}=\frac{1}{1.044\ 8}=0.957\ 1$$

这就是 $\boldsymbol{M}$ 矩阵的第一列.

用 $\lambda_2=-1.414\ 7$, 重复上述步骤, 求得 $m_{12}=-0.957\ 4$ 和 $m_{22}=0.288\ 8$, 即是 $\boldsymbol{M}$ 的第二列. 于是有

$$\boldsymbol{M}=\begin{bmatrix}0.289\ 7 & -0.957\ 4\\ 0.957\ 1 & 0.288\ 8\end{bmatrix}$$

$\boldsymbol{w}$ 和 $\boldsymbol{x}$ 之间的关系是

$$\begin{bmatrix}w_1\\ w_2\end{bmatrix}=\begin{bmatrix}0.289\ 7 & 0.957\ 1\\ -0.957\ 4 & 0.288\ 8\end{bmatrix}\begin{bmatrix}x_1-0.389\\ x_2-0.306\end{bmatrix}$$

或

$$w_1=0.289\ 7(x_1-0.389)+0.957\ 1(x_2-0.306)$$

$$w_2=-0.957\ 4(x_1-0.389)+0.288\ 8(x_2-0.306)$$

如果想研究稳定点附近的响应曲面, 我们就应该在 (w_1, w_2) 空间内确定合适的点, 在这些点上取观测值, 然后用上述的关系式将这些点变换为 (x_1, x_2) 空间中的点, 那样, 就可以进行试验了.

11.3.3 岭系统

11.3.2 节讨论了响应曲面的纯最大值点、最小值点、或鞍点, 我们也经常遇到它们的各种变形. 尤其是岭系统, 相当普遍. 考虑前面 (11.9) 式给出的二阶模型的正则形式:

$$\hat{y}=\hat{y}_s+\lambda_1w_1^2+\lambda_2w_2^2+\cdots+\lambda_kw_k^2$$

现在, 设稳定点 $\boldsymbol{x}_s$ 在实验的区域之内, 又设有一个或多个 λ_i 很小 (即 $\lambda_i\approx0$). 于是, 响应变量对系数 λ_i 小的变量 w_i 就很不灵敏了.

用于说明这种情况的等高线图显示在图 11.12 中, 其中 $k=2$ 个变量且 $\lambda_1=0$ (在实践中, λ_1 会是接近于零但不精确等于零). 此响应曲面的正则模型理论上是

$$\hat{y}=\hat{y}_s+\lambda_2w_2^2$$

其中 λ_2 是负的. 注意沿 w_1 方向拉长而成为一条 $\hat{y}=70$ 的中心线, 且在这一直线上的任意点处都取得最优值. 这类响应曲面就称为**稳定岭系统**.

如果稳定点远离二阶拟合模型的探测区域并且有一个 (或多个) λ_i 接近零, 则此曲面可能是一个**上升岭**. 图 11.13 说明一个上升岭, 此时 $k=2$ 个变量且 λ_1 接近零, λ_2 是负的. 在此类岭系统中, 我们不能进行有关真实曲面或稳定点的推断, 因为 $\boldsymbol{x}_s$ 在所拟合模型的区域之外, 所以应沿 w_1 方向作进一步探测. 如果 λ_2 是正的, 则称此系统为下落岭系统.

11.3.4 多重响应

许多响应曲面问题中包含了对几个响应的分析. 例如, 在例 11.2 中, 实验者测量了 3 个响应. 那个例子只对产率响应 y_1 优化了过程.

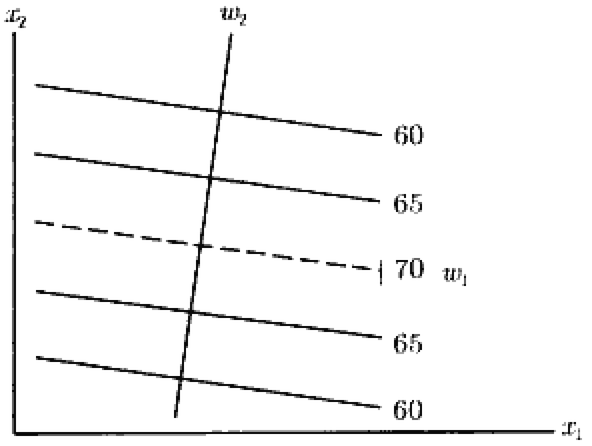


图 11.12 稳定岭系统的等高线图

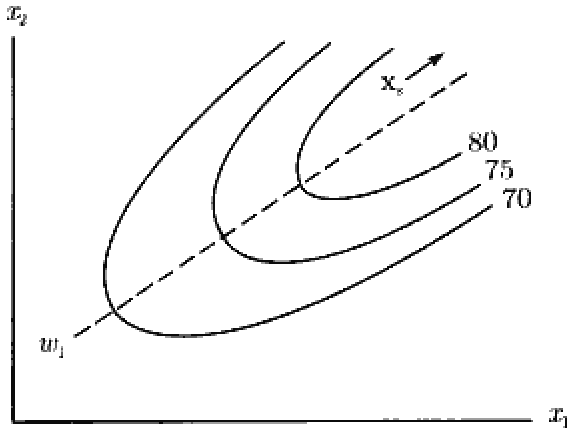


图 11.13 上升岭系统的等高线图

同时考虑多个响应的过程包括：首先，对每一个响应建立一个合适的响应曲面；然后，尝试找到一组运行条件，使得这些运行条件在某种意义下优化所有响应，或至少使各响应保持在理想范围内. 有关多重响应问题的更多的处理方法在 Myers and Montgomery (2002) 中给出.

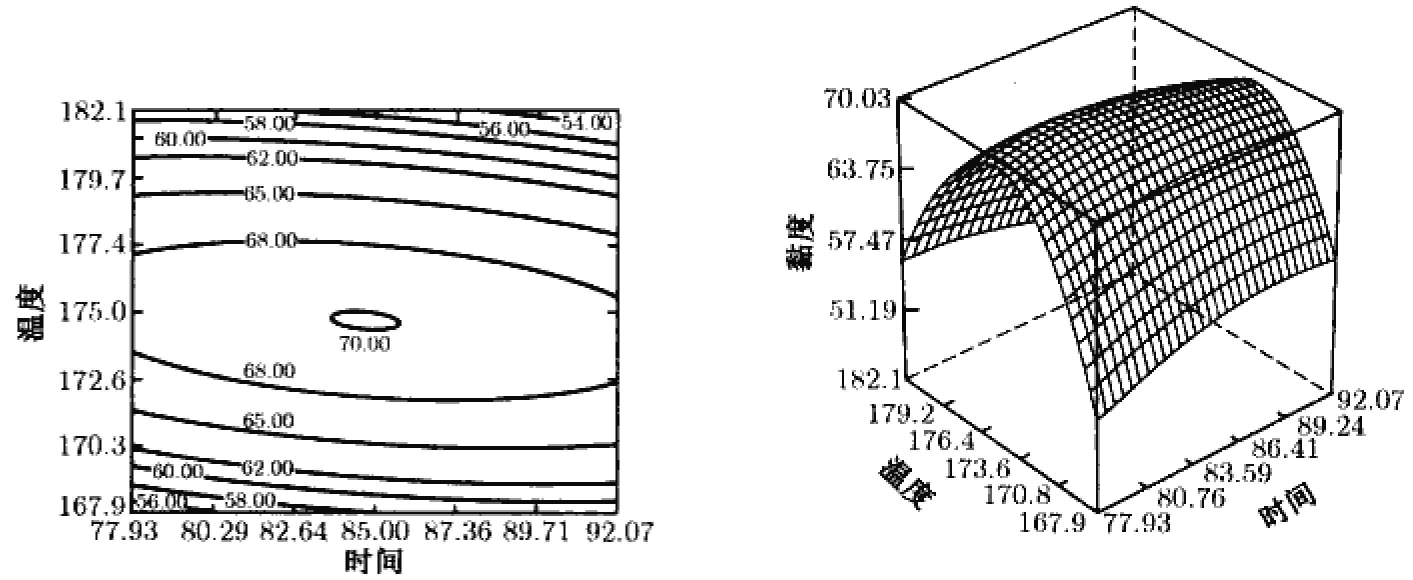
可以求得例 11.2 中黏度和分子量响应 (分别是 y_2 和 y_3) 的模型如下：

$$\hat{y}_2 = 70.00 - 0.16x_1 - 0.95x_2 - 0.69x_1^2 - 6.69x_2^2 - 1.25x_1x_2$$
$$\hat{y}_3 = 3\ 386.2 + 205.1x_1 + 177.4x_2$$

用时间 (ξ_1) 与温度 (ξ_2) 的自然水平写出上面两个模型，得

$$\hat{y}_2 = -9\ 030.74 + 13.393\xi_1 + 97.708\xi_2 - 2.75 \times 10^{-2}\xi_1^2 - 0.267\ 57\xi_2^2 - 5 \times 10^{-2}\xi_1\xi_2$$
$$\hat{y}_3 = -6\ 308.8 + 41.025\xi_1 + 35.473\xi_2$$

图 11.14 和图 11.15 显示了这两个模型的等高线图和响应曲面图.

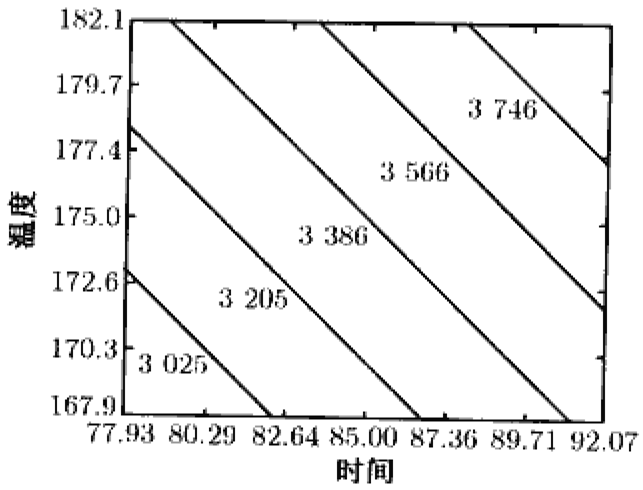


(a) 等高线图 (b) 响应曲面图

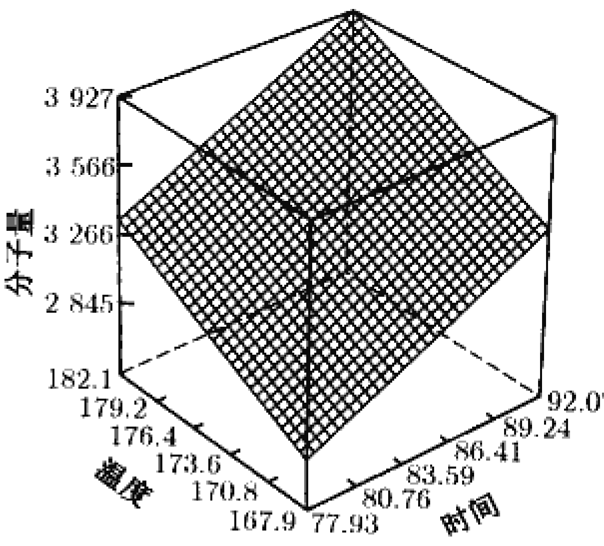
图 11.14 例 11.2 黏度响应的等高线图和响应曲面图

一个相对较直接的优化多个响应的方法是作出各响应的**重叠等高线图**，这种方法只在过程变量个数较少时效果较好. 图 11.16 显示了例 11.2 的 3 个响应的重叠等高线图，其中 y_1 (产率) $\geqslant 78.5$, $62 \leqslant y_2$ (黏度) $\leqslant 68$, y_3 (分子量) $\leqslant 3\ 400$ 的等高线. 如果这些边界表示过程必须满足的重要条件，那么图 11.16 中没有阴影的部分表示满足过程要求的时间和温度的一些组合. 实验

者可以目测等高线图以确定适当的运行条件. 例如, 实验者很可能对图 11.16 中两个可行运行区域中较大的一个感兴趣.



(a) 等高线图



(b) 响应曲面图

图 11.15　例 11.2 分子量响应的等高线图和响应曲面图

当存在 3 个以上的设计变量时, 重叠等高线图就变得难以使用了, 因为等高线图是二维的, 必须有 $k-2$ 个设计变量取常数才能作出图形. 为了获得曲面的最佳视野, 往往需要多次试错才能决定哪些因子应取常数以及它们应选取什么水平. 因此, 多重响应的正规最优化方法有着实际的需要.

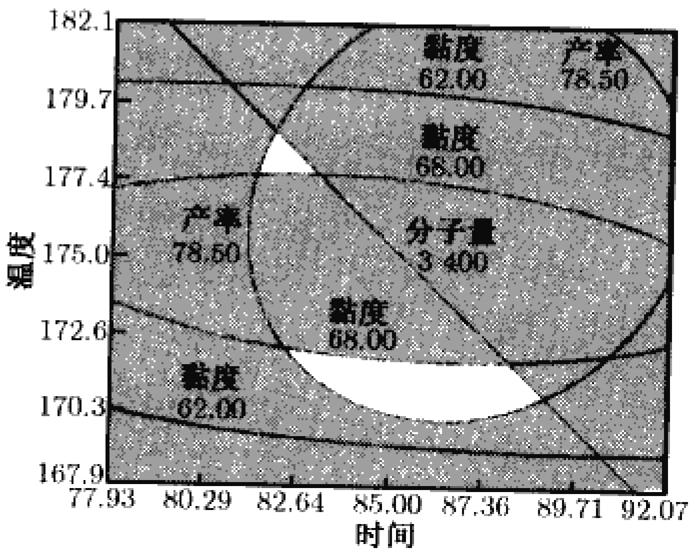


图 11.16　例 11.2 中用重叠产率、黏度和分子量等高线图找到的最优区域

一种常用的方法是像一个**约束最优化问题**那样, 列出方程并求解. 以例 11.2 说明, 我们对此问题可以列出如下方程:

$$\begin{aligned} &\text{Max} \quad y_1 \\ &\text{s.t.} \quad 62 \leqslant y_2 \leqslant 68 \\ &\qquad\;\; y_3 \leqslant 3\,400 \end{aligned}$$

许多数值方法可用于求解这个问题. 有时, 这些方法被称为**非线性规划法**. Design-Expert 软件包用直接搜索法求解这样的问题. 求出的两个解是

$$\text{时间} = 83.5 \qquad \text{温度} = 177.1 \qquad \hat{y}_1 = 79.5$$

和

$$时间 = 86.6 \qquad 温度 = 172.25 \qquad \hat{y}_1 = 79.5$$

注意到第一个解在设计空间 (参见图 11.16) 上面 (较小) 的可行域内, 而第二个解在较大的可行域内. 两个解都非常接近约束的边界.

另一种多重响应优化的方法是, 使用由 Derringer and Suich (1980) 推广的同时优化技术. 这种方法要利用**满意度函数**. 一般的方法是先将各个响应 y_i 转换为单个满意度函数 d_i, 其变化范围是

$$0 \leqslant d_i \leqslant 1$$

如果响应 y_i 是它的目标值, 则 $d_i= 1$; 如果响应在可接受的范围之外, 则 $d_i= 0$. 然后, 选择设计变量, 使之最大化 m 个响应的总满意度

$$D = (d_1 \cdot d_2 \cdot \cdots \cdot d_m)^{1/m}$$

单个满意度函数的构造方式见图 11.17. 如果响应 y 的目标 T 是一个最大值, 则

$$d = \begin{cases} 0 & y < L \\ \left(\frac{y-L}{T-L}\right)^r & L \leqslant y \leqslant T \\ 1 & y > T \end{cases} \tag{11.11}$$

当权重 $r = 1$ 时, 满意度函数是线性函数. 若选择 $r > 1$, 则更强调靠近目标值; 若选择 $0 < r < 1$, 则目标值较不重要. 如果响应的目标是一个最小值, 则

$$d = \begin{cases} 1 & y < T \\ \left(\frac{U-y}{U-T}\right)^r & T \leqslant y \leqslant U \\ 0 & y > U \end{cases} \tag{11.12}$$

双边的满意度函数显示在图 11.17c 中, 假定目标位于下限 (L) 和上限 (U) 中, 定义为

$$d = \begin{cases} 0 & y < L \\ \left(\frac{y-L}{T-L}\right)^{r_1} & L \leqslant y \leqslant T \\ \left(\frac{U-y}{U-T}\right)^{r_2} & T \leqslant y \leqslant U \\ 0 & y > U \end{cases} \tag{11.13}$$

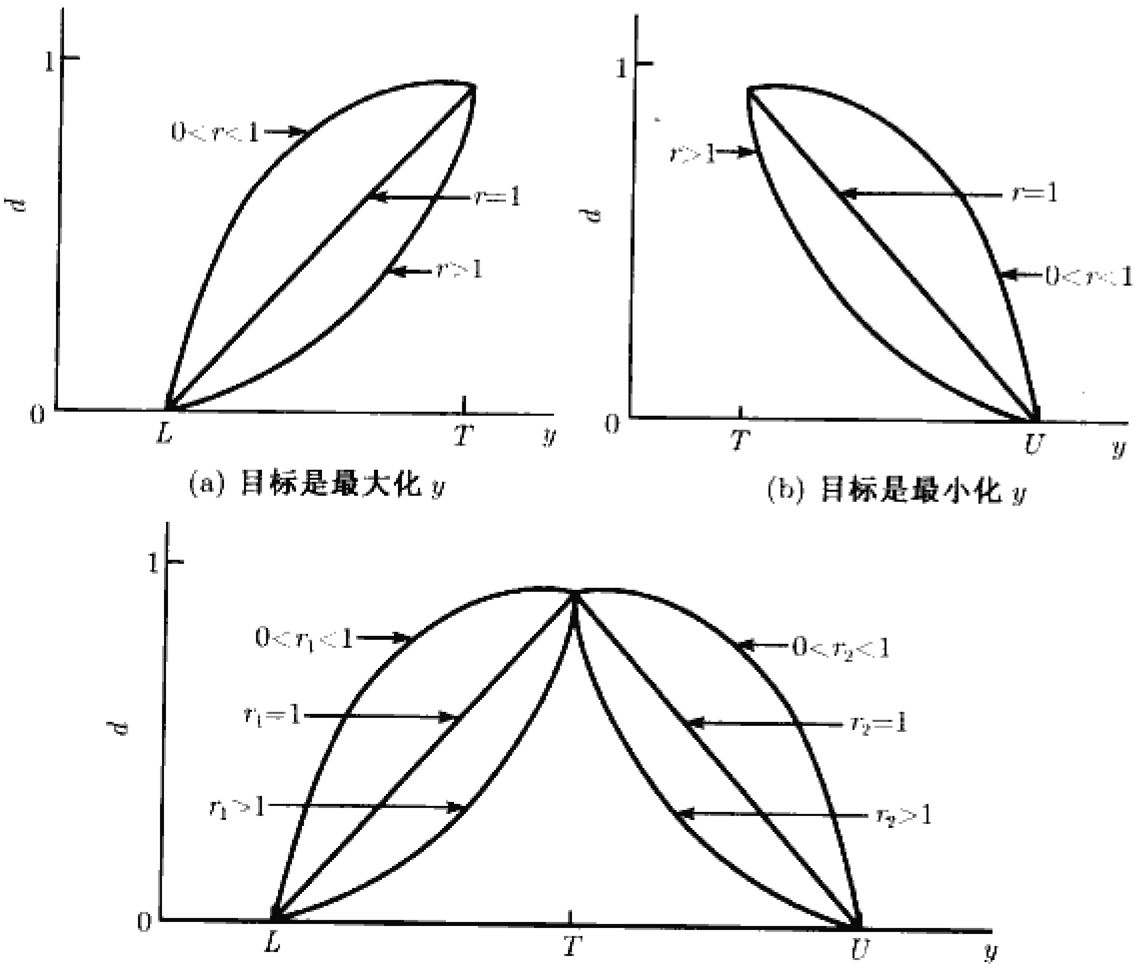
利用 Design-Expert 软件包的满意度函数法求解例 11.2. 选择 $T = 80$ 为产率响应的目标, $U = 70$, 并设单个满意度函数的权重是 1. 对黏度响应设 $T = 65$, $L = 62$, $U = 68$ (与规格限一致), 权重 $r_1 = r_2= 1$. 最后, 我们指出, 分子量在 3 200 到 3 400 之间是合适的. 求得两个解:

解 1: $\quad 时间 = 86.5 \quad 温度 = 170.5 \quad D = 0.822$

$\qquad\qquad \hat{y}_1 = 78.8 \quad \hat{y}_2 = 65 \quad \hat{y}_3 = 3\,287$

解 2: $\quad 时间 = 82 \quad 温度 = 178.8 \quad D = 0.792$

$\qquad\qquad \hat{y}_1 = 78.5 \quad \hat{y}_2 = 65 \quad \hat{y}_3 = 3\,400$

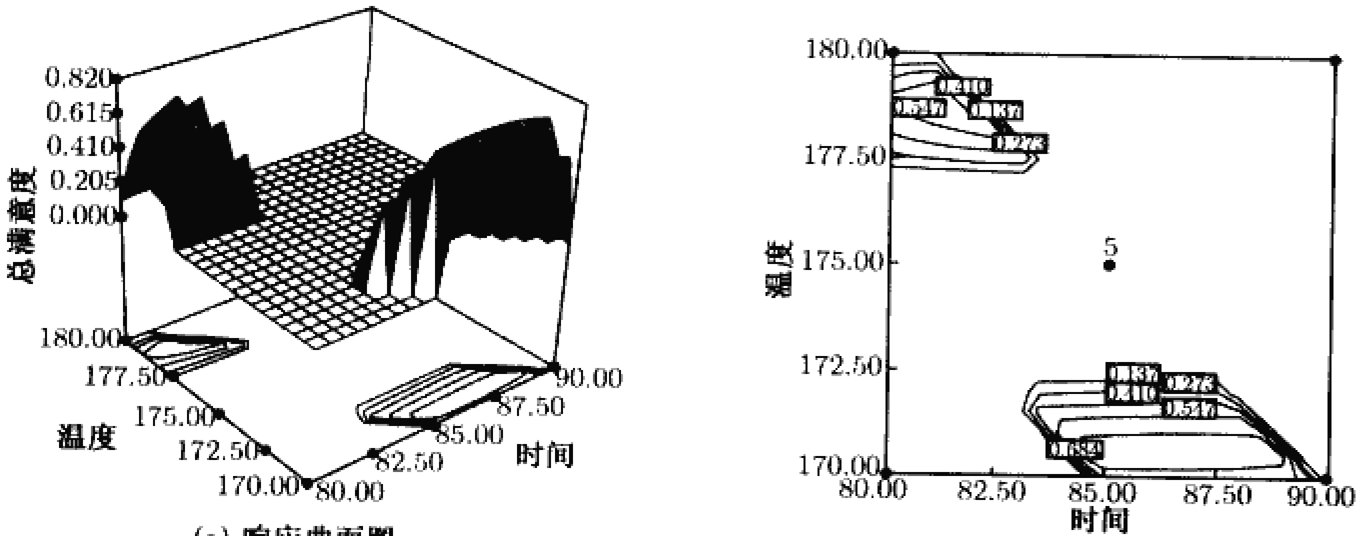


(a) 目标是最大化 y (b) 目标是最小化 y

(c) 目标是使 y 值尽可能靠近目标值

图 11.17 同时优化方法的单个满意度函数

解 1 有最高的总满意度值, 其黏度达到目标值, 分子量在可接受范围内. 这个解位于图 11.16 中两个运行区域中较大的一个区域中, 第 2 个解在较小的区域中. 图 11.18 显示了总满意度函数 D 的响应曲面和等高线图.



(a) 响应曲面图 (b) 等高线图

图 11.18 例 11.2 问题中满意度函数的响应曲面和等高线图

11.4 拟合响应曲面的实验设计

用适当选取的实验设计方法来拟合和分析响应曲面会带来极大的方便. 本节讨论为拟合响应曲面选择适当设计的某些方面.

在选择响应曲面的设计时, 理想的设计应具有下面一些特点.

(1) 在所研究的整个区域内, 能够提供数据点的合理分布以及其他信息.

(2) 容许研究模型的合适性, 包括拟合不足.

(3) 容许分区组进行实验.

(4) 容许逐步建立较高阶的设计.

(5) 提供内部的误差估计量.

(6) 提供模型系数的正确估计.

(7) 提供在实验区域内好的预测方差.

(8) 对异常值或缺失数据提供适当的稳健性.

(9) 不需要大量的试验.

(10) 不需要自变量有太多的水平.

(11) 确保模型参数计算的简单性.

这些特点有时是相互矛盾的, 所以, 在应用到设计选择中去时, 必须经常加以判断. 要得到关于选择响应曲面设计的更多信息, 请参阅 Myers and Montgomery (2002), Box and Draper (1987) 以及 Khuri and Cornell (1996).

11.4.1 拟合一阶模型的设计

设要拟合 k 个变量的一阶模型

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon \tag{11.14}$$

我们有一类独特的设计, 它使得回归系数$\{\hat{\beta}_i\}$ 的方差极小化. 这就是**正交一阶设计**. 一个一阶设计是正交的, 如果 $(\boldsymbol{X}'\boldsymbol{X})$ 矩阵的非对角元素全为零. 这一点意味着 $\boldsymbol{X}$ 矩阵列的叉积之和为零.

正交的一阶设计类包含了主效应不能互为别名的 2^k 析因设计和 2^k 分式析因设计系列. 在使用这些设计时, 我们假定 k 个因子的低水平与高水平被规范为通常的水平 ± 1.

2^k 设计不能提供实验误差的估计量, 除非某些试验重复进行. 使 2^k 设计包括有重复试验的常用方法是, 在设计的中心点 (点 $x_i = 0, i = 1, 2, \cdots, k$) 处增添几个观测值. 给 2^k 设计增加中心点并不影响$\{\hat{\beta}_i\}$, $i \geqslant 1$, 但是 β_0 的估计量将变为所有观测值的总平均值. 而且, 增加的中心点不影响设计的正交性. 例 11.1 说明了给 2^2 设计增添 5 个中心点来拟合一阶模型的用法.

另一种正交一阶设计是**单纯形**设计. 单纯形就是 k 维空间中有 $k+1$ 个顶点的等边图形. 这样一来, $k = 2$ 的单纯形设计是一个等边三角形, $k = 3$ 时是正四面体. 二维和三维的单纯形设计如图 11.19 所示.

11.4.2 拟合二阶模型的设计

我们在例 11.2 (甚至更早, 如例 6.6) 中已正式介绍了拟合二阶模型的**中心复合设计**(Central Composite Design, CCD). 这是用于拟合这些模型的最广的一类设计. 一般而言, CCD 是由有 n_F 个试验的 2^k 析因设计 (或分辨度为 V 的分式析因设计) 以及 $2k$ 个坐标轴试验点或星号点、n_C 个中心试验点组成的. 图 11.20 显示了 $k = 2$ 和 $k = 3$ 个因子的 CCD.

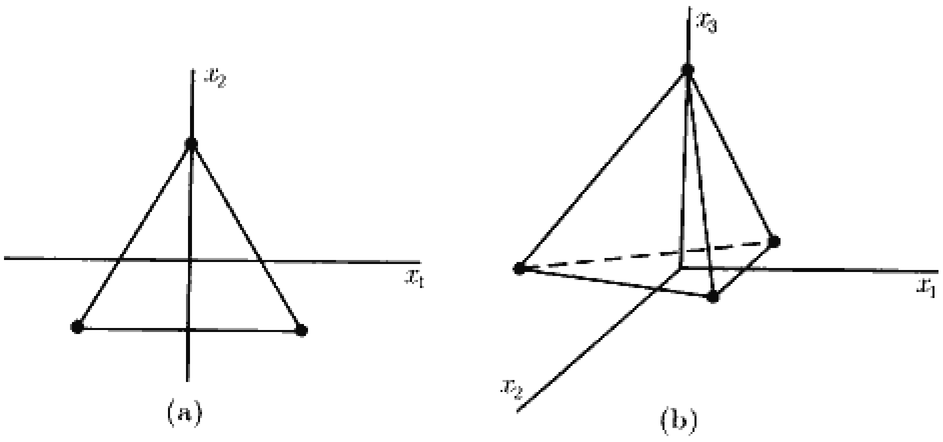


图 11.19 单纯形设计: (a)$k=2$ 个变量, (b) $k=3$ 个变量

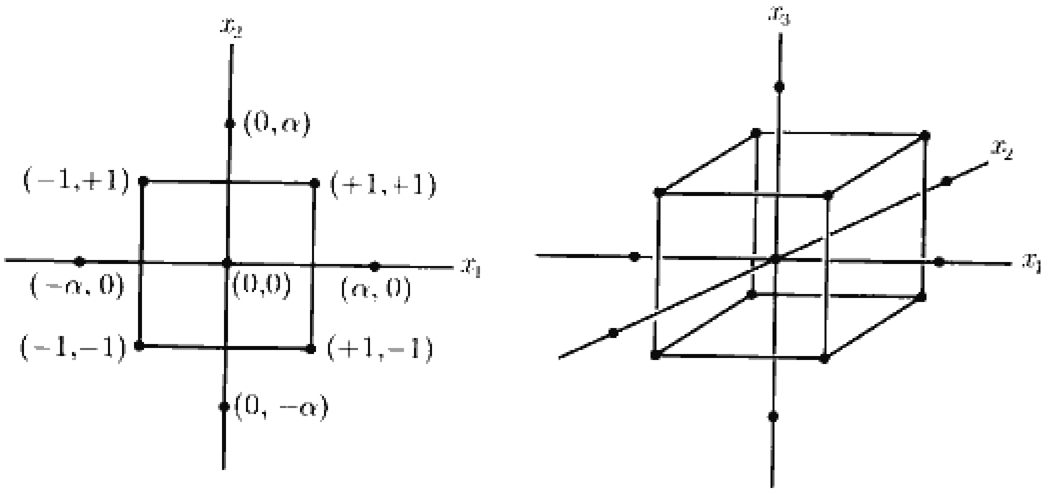


图 11.20 $k=2$ 和 $k=3$ 的中心复合设计

如例 11.1 和例 11.2 所示, CCD 的实际应用过程通常由**序贯实验**产生. 即, 2^k 设计已用于拟合一阶模型, 而这个模型显露出拟合不足, 然后增加坐标轴上的试验点, 使得二次项可以加入模型. CCD 对拟合二阶模型而言是非常有效的设计. 设计中必须指定两个参数: 从设计中心到坐标轴试验点的距离 α 和中心点个数 n_C. 以下讨论如何选取这两个参数.

1. **可旋转性**

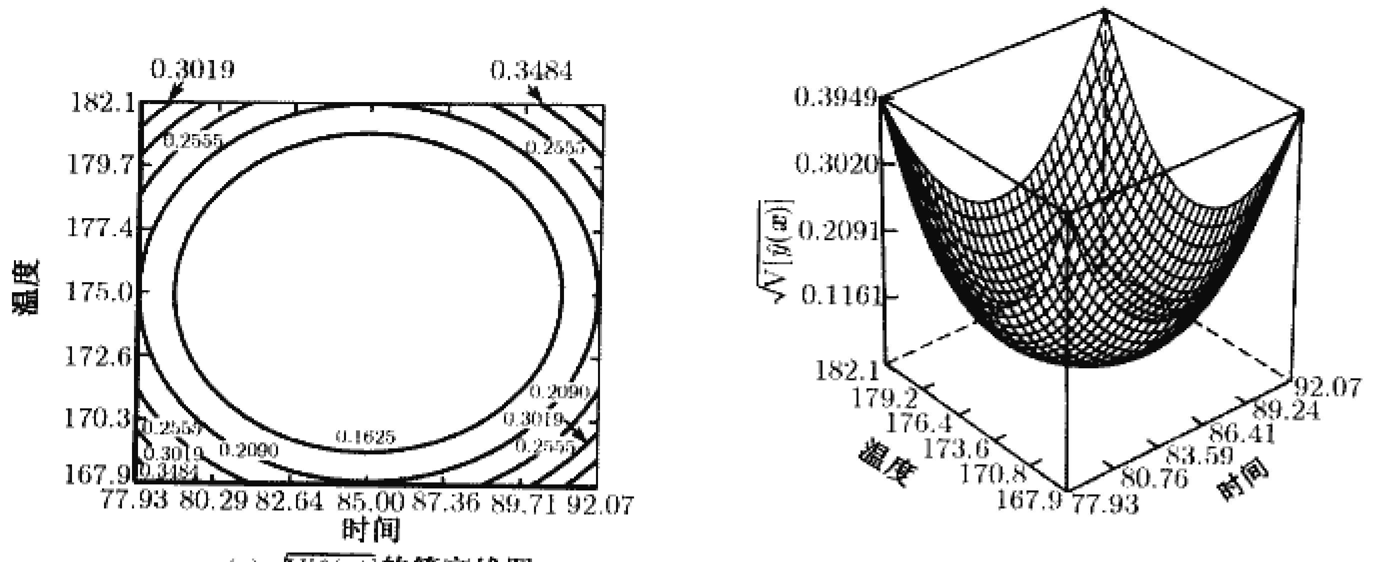
对二阶模型来说, 在所关注的整个区域内提供良好的预测是很重要的. 定义 "良好" 的一种方法是, 要求模型在关注点 $\boldsymbol{x}$ 处的预测响应有相当一致和稳定的方差. 由 (10.40) 式知, 在点 $\boldsymbol{x}$ 处预测响应的方差是

$$\mathrm{V}[\hat{y}(\boldsymbol{x})]=\sigma^2\boldsymbol{x}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x} \tag{11.15}$$

Box and Hunter (1957) 建议, 二阶响应曲面设计应该是**可旋转的**. 这意味着在所有与设计中心距离相同的 $\boldsymbol{x}$ 点处 $\mathrm{V}[\hat{y}(\boldsymbol{x})]$ 相等. 即, 预测响应的方差在球面上是常数.

图 11.21 显示了在例 11.2 中用 CCD 的二阶模型拟合的 $\sqrt{\mathrm{V}[\hat{y}(\boldsymbol{x})]}$ 为常数的等高线. 注意到预测响应的标准差为常数的等高线是同心圆. 具有这一性质的设计, 当它围绕中心点 $(0, 0, \cdots, 0)$ 旋转时, 将会使 $\hat{y}$ 的方差保持不变, 因而称为*可旋转*设计.

对选择响应曲面设计而言, 可旋转性是一个合理的依据. 因为 RSM 的目的是优化, 而最优点的位置在做实验前是未知的, 可旋转性的意义就在于所使用的设计在各个方向上提供等精确度的估计. (可以证明任意一阶正交设计是可旋转的.)



(a) $\sqrt{V[\hat{y}(x)]}$的等高线图　　(b) 响应曲面图

图 11.21　例 11.2 可旋转 CCD 中预测响应的标准差的等高线

恰当选择 α 可以使中心复合设计是可旋转的. 满足可旋转性要求的 α 值依赖于设计的析因部分内点的数目; 事实上, $\alpha = (n_F)^{1/4}$ 能产生可旋转中心复合设计, 其中 n_F 是用在设计的析因部分中的点的个数.

2. 球面 CCD

可旋转性是一种球面性质, 即, 当所关注的区域是球面时, 可旋转性是最有意义的设计准则. 然而, 严格的可旋转性对于一个良好的设计而言并不重要. 对于球形区域, 从 CCD 的预测方差的观点来看, α 的最好的选择是设 $\alpha = \sqrt{k}$. 将所有的析因设计点和坐标轴上的设计点都放在半径为 $\sqrt{k}$ 的球体表面的设计称为**球面 CCD**. 对此更详细的讨论见 Myers and Montgomery (2002).

3. CCD 的中心试验点

在 CCD 中, α 的选择主要由感兴趣的区域决定. 当这个区域是球面时, 设计必定包含中心试验点以提供预测响应的合理稳定方差. 一般推荐用 3~5 个中心试验点.

4. Box-Behnken 设计

Box and Behnken (1960) 提出过一些拟合响应曲面的三水平设计. 这些设计由 2^k 析因设计与不完全区组设计组合而成. 所得出的设计对所要求做的试验次数来说, 十分有效, 而且它们是可旋转的或接近可旋转的.

表 11.8 列出了 3 个变量的 Box-Behnken 设计. 图 11.22 是此设计的图解. 注意到 Box-Behnken 设计是一个球面设计, 所有设计点都在半径为 $\sqrt{2}$ 的球面上. 而且 Box-Behnken 设计不包含由各个变量的上限和下限所生成的立方体区域的顶点处的任一点. 当立方体顶点所代表的因子水平组合因试验成本过于昂贵或因实际限制而不可能做试验时, 此设计就显示出它特有的长处.

5. 立方体形区域

在许多情况下, 所关注的区域是立方体形而不是球形的. 此时, **面中心的中心复合设计**或称**面中心立方**是中心复合设计的一种变形, 其中 $\alpha = 1$. 此设计在各坐标轴上所取的点是立方体各

个面上的中心点, 对 $k=3$, 如图 11.23 所示. 有时采用中心复合设计的这种变形, 因为它只要求每个因子有 3 个水平, 而在实践中, 经常难于改变因子水平. 但是, 要注意面中心的中心复合设计是不可旋转的.

表 11.8　3 个变量的 Box-Behnken 设计

试验	x_1	x_2	x_3	试验	x_1	x_2	x_3
1	−1	−1	0	9	0	−1	−1
2	−1	1	0	10	0	−1	1
3	1	−1	0	11	0	1	−1
4	1	1	0	12	0	1	1
5	−1	0	−1	13	0	0	0
6	−1	0	1	14	0	0	0
7	1	0	−1	15	0	0	0
8	1	0	1				

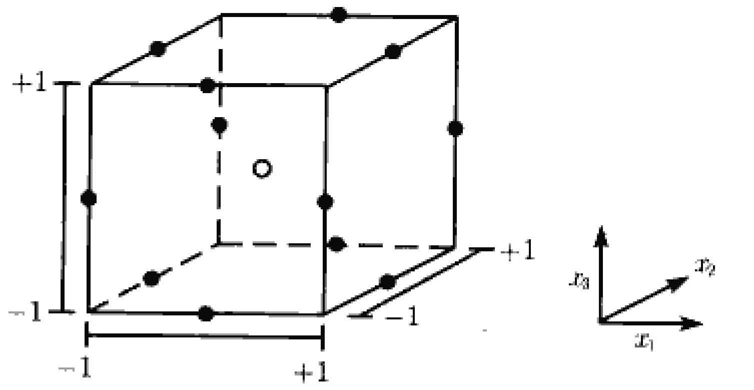


图 11.22　三因子的 Box-Behnken 设计

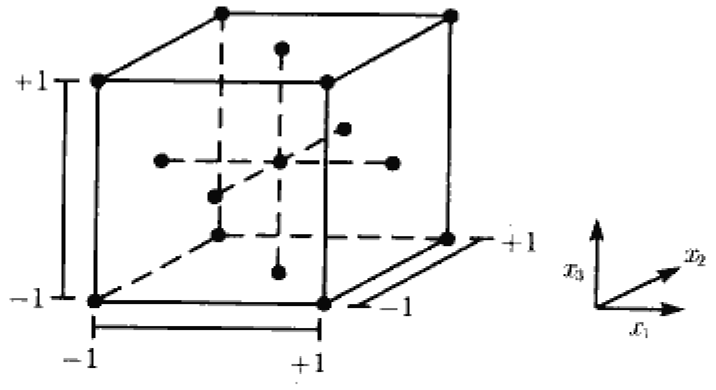
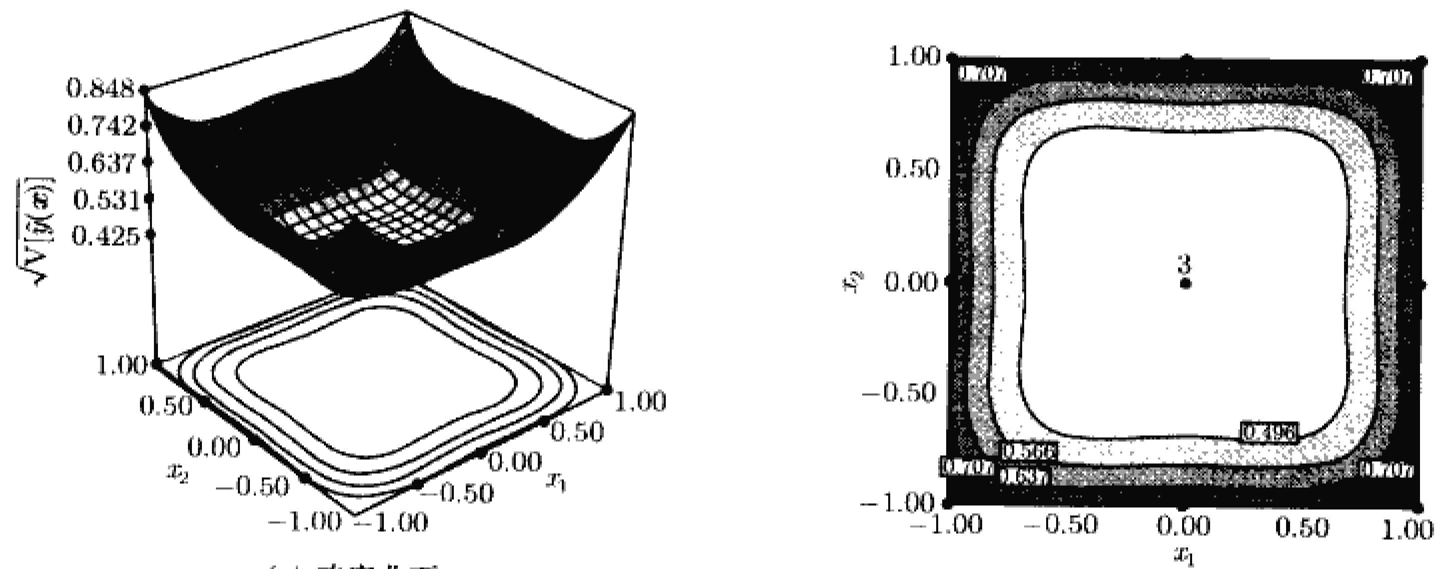


图 11.23　$k=3$ 的面中心的中心复合设计

面中心立方并不像球形 CCD 那样需要许多中心点. 实际上, $n_C=2$ 或 3 就足以对整个实验区域提供好的预测方差. 应该注意, 有时用更多的中心试验点可以给出实验误差更合理的估计. 图 11.24 显示了 $k=3$ 且 $n_C=3$ 个中心点 ($x_3=0$) 时, 面中心立方的预测方差平方根 $\sqrt{V[\hat{y}(\boldsymbol{x})]}$ 的相关图形. 预测响应的标准差在设计空间相对较大的区域内是相当一致的.



(a) 响应曲面　(b) 等高线图

图 11.24　$k=3, n_C=3, x_3=0$ 的面中心立方的预测响应的标准差 $\sqrt{V[\hat{y}(\boldsymbol{x})]}$

6. 其他设计

在实践中偶尔也使用许多其他的响应曲面设计. 对于两个变量, 可以用这样的设计, 它由圆周上等距离的点所组成, 并构成正多边形. 因为设计点与原点是等距离的, 所以, 这些设计常称为**等半径设计**.

$k=2$ 时, 可旋转等半径设计可以在圆周上取 $n_2 \geqslant 5$ 个等距离的点并在圆心上取 $n_1 \geqslant 1$ 个点组合起来而求得. $k=2$ 时, 特别有用的设计是五边形和六边形. 这些设计如图 11.25 所示. 其他有用的设计包括**小复合设计**和**混杂设计类**. 小复合设计由分辨度为 III*(主效应与二因子交互作用混杂, 二因子交互作用之间无混杂) 的立方体中的分式析因设计, 以及通常的坐标轴点和中心点构成. 在需要尽可能减少试验次数时, 可以考虑这两种设计.

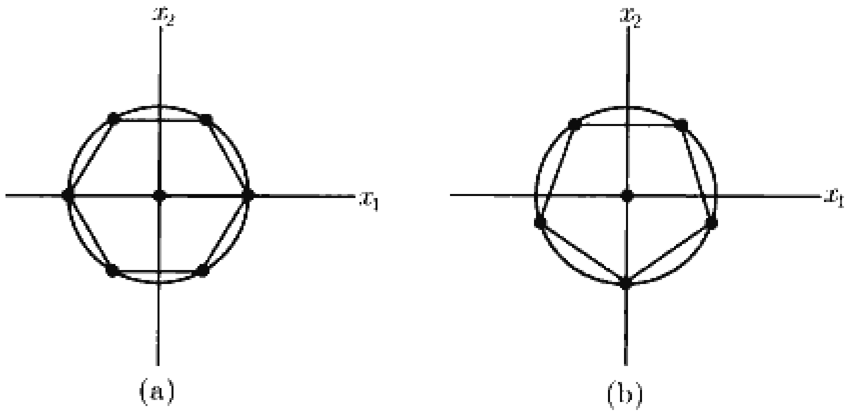


图 11.25 两个变量的等半径设计: (a) 正六边形 (b) 正五边形

$k=3$ 个因子的小复合设计如表 11.9 所示. 这个设计在立方体中用 2^3 设计的标准 1/2 分式设计, 因为它满足分辨度为 III* 的要求. 此设计有 4 次立方体上的试验, 6 次坐标轴上的试验, 且必须至少有一次中心点上的试验. 于是, 该设计至少有 $N=11$ 次试验, $k=3$ 个变量的二阶模型要估计 $p=10$ 个参数, 所以对试验次数而言是非常有效的设计. 表 11.9 中的设计有 n_C=4 个中心点. 我们选择 α=1.73 以得到球形设计, 因为小复合设计不可能是可旋转的设计.

表 11.9 $k=3$ 个因子的小复合设计

标准次序	x_1	x_2	x_3	标准次序	x_1	x_2	x_3
1	1.00	1.00	−1.00	8	0.00	1.73	0.00
2	1.00	−1.00	1.00	9	0.00	0.00	−1.73
3	−1.00	1.00	1.00	10	0.00	0.00	1.73
4	−1.00	−1.00	−1.00	11	0.00	0.00	0.00
5	−1.73	0.00	0.00	12	0.00	0.00	0.00
6	1.73	0.00	0.00	13	0.00	0.00	0.00
7	0.00	−1.73	0.00	14	0.00	0.00	0.00

表 11.10 列出了 $k=3$ 时的混杂设计. 这类设计中的一些设计有不规则的水平, 在应用时会受到限制. 然而, 它们是很小的设计, 并有优良的预测方差的性质. 有关小复合设计和混杂设计更详细的内容, 可参阅 Myers and Montgomery (2002).

7. 响应曲面设计的图形评估

响应曲面设计最常用于为了预测而建立模型. 因此, [由 (11.15) 式定义的] 预测方差在评估或比较设计时是相当重要的. 类似于图 11.21 和图 11.24, 预测方差 (或它的平方根, 预测标准

差) 的二维等高线图或三维响应曲面图在这方面是有价值的. 但对于有 k 个因子的设计, 这些图形只允许两个设计因子显示在图中. 因为所有其余 $k-2$ 个因子在图中为常数, 所以这些图对预测方差在整个设计空间如何分布给出的是不完整的图像. Giovannitti-Jensen and Myers (1989) 为了解决这个问题提出了**方差散布图** (VDG).

表 11.10 $k = 3$ 个因子的混杂设计

标准次序	x_1	x_2	x_3	标准次序	x_1	x_2	x_3
1	0.00	0.00	1.41	7	1.41	0.00	−0.71
2	0.00	0.00	−1.41	8	−1.41	0.00	−0.71
3	−1.00	−1.00	0.71	9	0.00	1.41	−0.71
4	1.00	−1.00	0.71	10	0.00	−1.41	−0.71
5	−1.00	1.00	0.71	11	0.00	0.00	0.00
6	1.00	1.00	0.71				

VDG 对特定的设计和响应模型, 以图形方式显示了到中心点有相同距离的设计点的预测方差的最小值、最大值和平均值. 距离或半径通常在零 (设计中心) 到 $\sqrt{k}$ 之间变化, 对球形设计, $\sqrt{k}$ 是设计点到中心点的最大距离. 在 VDG 中通常以**尺度化预测方差**(SPV)

$$\frac{NV[\hat{y}(\boldsymbol{x})]}{\sigma^2} = N\boldsymbol{x}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x} \tag{11.16}$$

作图. 注意, SPV 是 (11.15) 式中的预测方差乘以设计试验次数 (N) 再除以误差的方差 σ^2. 除以 σ^2 消除了一个未知参数, 而乘以 N 便于比较试验次数不同的设计.

图 11.26a 是有 $k = 3$ 个变量、4 个中心试验点的可旋转 CCD 的 VDG. 因为此设计是可旋转的, 所有到设计中心距离相同的点的最小、最大、平均 SPV 均相同, 所以在 VDG 中只有一条曲线. 图形显示了在整个设计空间中 SPV 的变化情况, 在半径小于约 1.2 时, SPV 近似为常数, 然后直到边界为止一直稳定增大. 图 11.26b 是有 $k = 3$ 个变量、4 个中心试验点的球形 CCD 的 VDG. 注意到最大、最小、平均 SPV 三条曲线只有很小的差异, 因此可以得到这样的结论: 可旋转设计和球形设计之间的实际差异非常小.

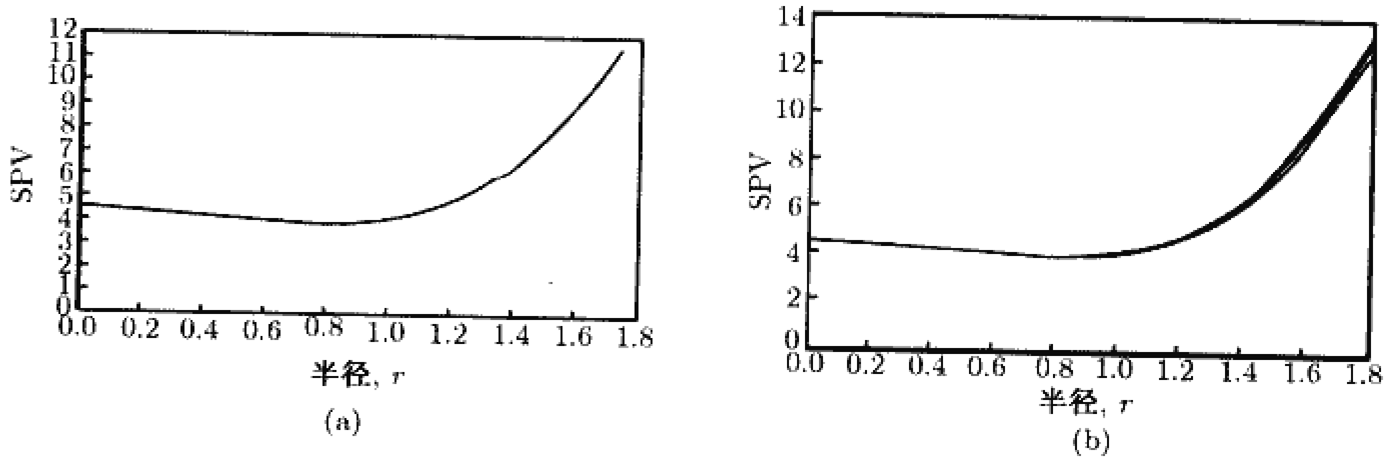


图 11.26 方差散布图: (a) $k = 3, \alpha = 1.68$(4 个中心点) 的 CCD
(b) $k = 3, \alpha = 1.732$(4 个中心点) 的 CCD

图 11.27 是有 $k = 4$ 个变量的可旋转 CCD 的 VDG. 在这个 VDG 中, 设计所含的中心

试验点个数从 $n_C = 1$ 到 $n_C = 5$ 变化. VDG 清楚地表明, 中心点个数太少的设计的预测方差有非常不稳定的分布, 但随着 n_C 的增加, 预测方差迅速趋于稳定. 用 4 个或 5 个中心试验点时, 在整个设计区域内会有相当稳定的预测方差. VDG 已被用于研究响应曲面设计中中心试验点个数的变化引起的效应. 本章前面的建议就是基于这些研究的.

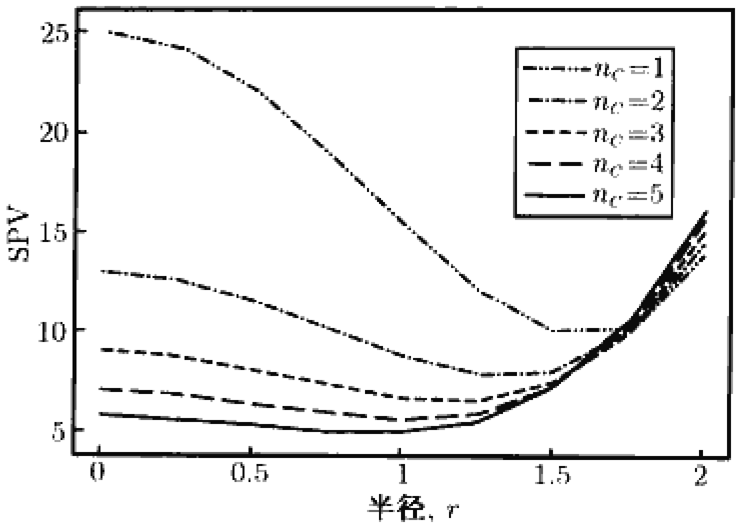


图 11.27 $k = 4, \alpha = 2$ 时 CCD 的方差散布图

11.4.3 响应曲面设计的区组化

当使用响应曲面设计时, 经常需要考虑划分区组, 以便消除多余的变量. 例如, 像在例 11.1 和例 11.2 中说明过的那样, 当二阶设计是由一阶设计序贯组装而成时, 就会出现这种问题. 在做一阶设计的试验和做建立二阶设计所需的增补试验两者之间, 间隔时间过长, 试验的条件可能就改变了. 因此, 划分区组就成为必要的了.

响应曲面设计称为**正交区组化**, 如果它划分成区组, 使得区组效应不影响响应曲面模型的参数估计. 当 2^k 或 2^{k-p} 设计用作一阶响应曲面设计时, 第 7 章的方法可用来将试验安排在 2^r 个区组内进行. 这些设计的中心点应该在各区组内同等地配置.

划分为正交区组的二阶设计必须满足两个条件. 如果在第 b 个区组中有 n_b 个观测值, 则这些条件如下.

(1) 每个区组必须是一阶正交设计, 也就是

$$\sum_{u=1}^{n_b} x_{iu}x_{ju} = 0 \qquad i \neq j = 0, 1, \cdots, k, \text{ 对一切 } b$$

其中 x_{iu} 和 x_{ju} 是实验的第 u 个试验中第 i 个和第 j 个变量的水平, 对所有 u 定义 $x_{0u} = 1$.

(2) 每个区组所贡献的每个变量的总平方和所占的比例必须等于出现在此区组中的总观测数所占的比例, 也就是

$$\frac{\sum\limits_{u=1}^{n_b} x_{iu}^2}{\sum\limits_{u=1}^{N} x_{iu}^2} = \frac{n_b}{N} \qquad i = 1, 2, \cdots, k, \text{ 对一切 } b$$

其中 N 是设计的试验总数.

作为应用这些条件的一个例子，考虑有 $N=12$ 个试验、$k=2$ 个变量的可旋转中心复合设计. 可将此设计的 x_1 和 x_2 的水平以设计矩阵的形式写为

$$
D = \begin{bmatrix} x_1 & x_2 \\ -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1.414 & 0 \\ -1.414 & 0 \\ 0 & 1.414 \\ 0 & -1.414 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}
\begin{matrix} \\ \left.\begin{matrix} \\ \\ \\ \\ \\ \\ \end{matrix}\right\} \text{区组 1} \\ \left.\begin{matrix} \\ \\ \\ \\ \\ \\ \end{matrix}\right\} \text{区组 2} \end{matrix}
$$

此设计分为两个区组，第一个区组由设计的析因部分加上两个中心点组成，第二个区组由坐标轴点加上两个附加的中心点组成. 显然符合条件 1，也就是说，两个区组都是一阶正交设计. 再看条件 2，考虑区组 1，注意到

$$\sum_{u=1}^{n_1} x_{1u}^2 = \sum_{u=1}^{n_1} x_{2u}^2 = 4$$

$$\sum_{u=1}^{N} x_{1u}^2 = \sum_{u=1}^{N} x_{2u}^2 = 8 \quad 以及\ n_1 = 6$$

因此，

$$\frac{\sum_{u=1}^{n_1} x_{iu}^2}{\sum_{u=1}^{N} x_{iu}^2} = \frac{n_1}{N}$$

即

$$\frac{4}{8} = \frac{6}{12}$$

于是区组 1 满足条件 2. 对区组 2，有

$$\sum_{u=1}^{n_2} x_{1u}^2 = \sum_{u=1}^{n_2} x_{2u}^2 = 4 \quad 以及\ n_2 = 6$$

因此，

$$\frac{\sum_{u=1}^{n_2} x_{iu}^2}{\sum_{u=1}^{N} x_{iu}^2} = \frac{n_2}{N}$$

即

$$\frac{4}{8} = \frac{6}{12}$$

因为区组 2 也满足条件 2, 因此, 此设计的区组是正交的.

一般说来, 中心复合设计总可以分为两个正交的区组, 第 1 个区组由 n_F 个析因点加 n_{CF} 个中心点组成, 第 2 个区组由 $n_A = 2k$ 个坐标轴点加 n_{CA} 个中心点组成. 不论设计中所用的 α 值是什么, 正交区组所要求的第 1 个条件总能成立. 要使第 2 个条件成立, 就要有

$$\frac{\sum\limits_{u}^{n_2} x_{iu}^2}{\sum\limits_{u}^{n_1} x_{iu}^2} = \frac{n_A + n_{CA}}{n_F + n_{CF}} \tag{11.17}$$

(11.17) 式的左边是 $2\alpha^{\,2}/n_F$, 将此量代入 (11.17) 式就解得正交区组的 α 值为

$$\alpha = \left[\frac{n_F(n_A + n_{CA})}{2(n_F + n_{CF})}\right]^{1/2} \tag{11.18}$$

一般说来, 此 α 值得不到可旋转设计或球形设计. 如果设计还要求是可旋转的, 则 $\alpha = (n_F)^{1/4}$, 并有

$$(n_F)^{1/2} = \frac{n_F(n_A + n_{CA})}{2(n_F + n_{CF})} \tag{11.19}$$

要求出正好满足 (11.19) 式的设计通常是不可能的. 例如, 当 $k = 3$ 时, 则 $n_F = 8$ 和 $n_A = 6$, (11.19) 式为

$$(8)^{1/2} = \frac{8(6 + n_{CA})}{2(8 + n_{CF})}$$

$$2.83 = \frac{48 + 8n_{CA}}{16 + 2n_{CF}}$$

不可能求出精确满足最后一个等式的 n_{CA} 和 n_{CF}. 但是, 如果 $n_{CF}= 3$, $n_{CA}= 2$, 则右边是

$$\frac{48 + 8(2)}{16 + 2(3)} = 2.91$$

所以, 设计的区组是近似正交的. 在实践中, 只要不损失重要的信息, 对可旋转性或区组的正交性的要求可以放宽一些.

中心复合设计在容纳区组化的能力方面是十分强大的. 如果 k 足够大, 设计的析因部分可以分为两个或多个区组 (其个数必须是 2 的幂, 而坐标轴点形成一个单一的区组). 表 11.11 介绍几个有用的对中心复合设计的区组化安排.

当响应曲面设计分组进行试验时, 方差分析有两个要点. 第一, 关于用来计算纯误差估计量的中心点的用法. 只有在同一个区组内进行试验的那些中心点才能看作为重复试验, 所以, 纯误差项只能在各个区组内计算, 如果变异性在各个区之间是一致的, 则这些纯误差可以合并起来. 第二, 关于区组效应. 如果设计分为 m 个正交的区组, 则区组平方和是

$$SS_{\text{区组}} = \sum_{b=1}^{m} \frac{B_b^2}{n_b} - \frac{G^2}{N} \tag{11.20}$$

其中 B_b 是第 b 个区组的 n_b 个观测值的总和, G 是 m 个区组的全体 N 个观测值的总和. 当区组不是精确地正交时, 可以用第 10 章描述的一般回归显著性检验法 (“附加平方和” 法).

表 11.11 一些划分为正交区组的可旋转的及近似可旋转的中心复合设计

k	2	3	4	5	5 $\frac{1}{2}$ 重复	6	6 $\frac{1}{2}$ 重复	7	7 $\frac{1}{2}$ 重复
析因区组									
n_F	4	8	16	32	16	64	32	128	64
区组数	1	2	2	4	1	8	2	16	8
每区组的点数	4	4	8	8	16	8	16	8	8
每区组的中心点数	3	2	2	2	6	1	4	1	1
每区组的总点数	7	6	10	10	22	9	20	9	9
坐标轴点区组									
n_A	4	6	8	10	10	12	12	14	14
n_{CA}	3	2	2	4	1	6	2	11	4
坐标轴点区组的总点数	7	8	10	14	11	18	14	25	18
设计的总点数 N	14	20	30	54	33	90	54	169	90
α值									
正交区组	1.414 2	1.633 0	2.000 0	2.366 4	2.000 0	2.828 4	2.366 4	3.333 3	2.828 4
可旋转性	1.414 2	1.681 8	2.000 0	2.378 4	2.000 0	2.828 4	2.378 4	3.363 6	2.828 4

11.4.4 计算机生成 (最优) 设计

前面各节所讨论的标准响应曲面设计, 如中心复合设计、Box-Behnken 设计, 以及它们的变形 (如面中心立方), 被广泛使用, 因为它们是相当通用的且灵活的设计. 如果实验区域是立方体或球体, 通常应用标准响应曲面设计来解决问题. 然而实验者有时会遇到不易选择标准响应曲面的情况. 这时**计算机生成**设计可用于这种情况.

有 3 种情况适合使用某种计算机生成设计.

(1) 不规则的实验区域. 如果实验区域不是立方体或球体, 标准设计就不是最好的选择. 经常会出现不规则区域. 例如, 实验者正在研究一种特殊粘合剂的性质. 把粘合剂涂在两个零件上, 然后高温烘干. 考虑的两个因子是粘合剂的用量和烘干温度. 这两个因子的全部取值范围在通常的规范变量尺度下是 -1 到 1, 实验者知道如果粘合剂用量太少且烘干温度太低, 零件就不会粘合得很好. 用规范变量写出的对设计变量的**约束**条件为

$$-1.5 \leqslant x_1 + x_2$$

其中 x_1 表示粘合剂的用量, x_2 表示温度. 此外, 如果温度太高而粘合剂太多的话, 零件或者会受到热应力的损害或者会导致粘合不充分. 于是, 对因子水平有另一个约束

$$x_1 + x_2 \leqslant 1$$

图 11.28 显示了具有这两个约束的实验区域. 约束移去了正方形的两个角, 产生了不规则的实验区域 (有时, 不规则区域称为 "瘪罐"). 在这个区域上不存在可以进行精确拟合的标准响应曲面设计.

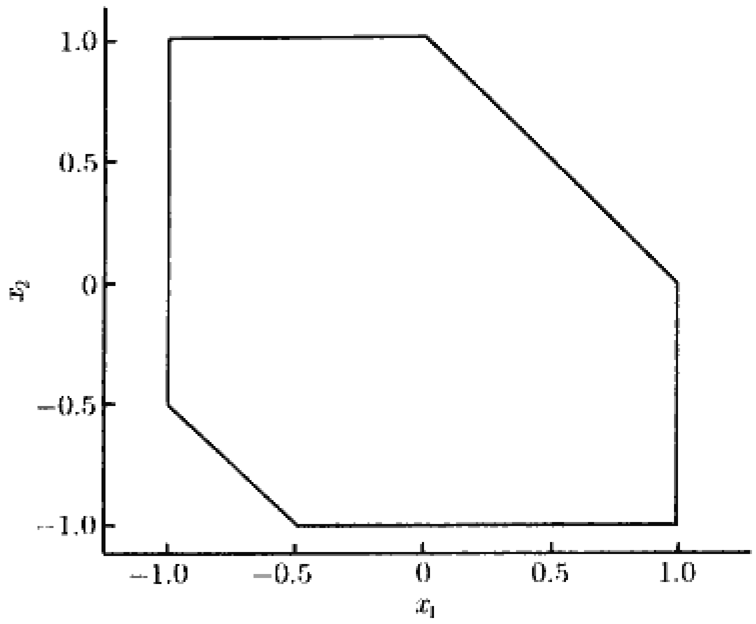


图 11.28 两个变量时的一个约束设计区域

(2) 非标准模型. 实验者一般选择一阶或二阶响应曲面模型作为对未知真实机理近似的**经验模型**. 然而有时实验者对所研究的实验有一些特殊的看法, 这时需要非标准模型. 例如, 需要模型

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_{12}x_1x_2+\beta_{11}x_1^2+\beta_{22}x_2^2+\beta_{112}x_1^2x_2+\beta_{1112}x_1^3x_2+\varepsilon$$

实验者想要获得拟合这个简化了的 4 次模型的一个有效的设计. 有时我们还会遇到部分因子是分类变量的响应曲面问题. 在这样的情况下, 不存在标准的响应曲面设计. [对于有分类变量的响应曲面问题, 参阅 Myers and Montgomery (2002)].

(3) 不寻常的样本量要求. 实验者偶尔会要求减少标准响应曲面设计所要求的试验次数. 例如, 要拟合有 4 个变量的二阶模型. 这种情况下的中心复合设计要求做 28 到 30 次试验, 具体试验次数由所选的中心点个数决定. 而模型只有 15 项, 如果试验成本极大, 或时间极长, 实验者会想用试验次数较少的设计. 尽管这种情况可以用计算机生成设计, 但也有更好的常用方法. 如 4 个因子的小复合设计有 20 次试验, 其中包含 4 个中心点; 混杂设计可以少到只有 16 次试验. 用计算机生成设计来减少试验次数一般是较好的选择.

计算机生成设计的许多进展是 Kiefer (1959,1961) 及 Kiefer and Wolfowitz (1959) 有关**最优设计**理论工作的推广结果. 所谓最优设计, 意味着它是关于某个准则的 "最好" 设计. 需要计算机程序来构造这些设计. 通常的方法是指定一个模型, 确定所在区域, 选择要做的试验次数, 指定最优化的准则, 然后从实验者考虑使用的*候选点*中选择设计点. 典型的候选点是分布在可行设计区域中的网格点.

有几个常用的设计最优化准则. 或许最常用的准则是 D 最优准则. 如果一个设计能最小化

$$|(\boldsymbol{X}'\boldsymbol{X})^{-1}|$$

则称之为**D 最优**. D 最优设计最小化了回归系数的联合置信区域. 在 D 最优准则下, 设计 1 关于设计 2 的相对效率定义为

$$D_e=\left(\frac{|(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}|}{|(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}|}\right)^{1/p} \tag{11.21}$$

其中, $\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$ 是两个设计的 $\boldsymbol{X}$ 矩阵, p 是模型中参数的个数.

A 最优准则只涉及回归系数的方差. 一个设计称为 A 最优设计, 如果它能最小化 $(\boldsymbol{X}'\boldsymbol{X}^{-1})$ 主对角元素的和 [称为 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 的**迹**, 记为 $\mathrm{tr}((\boldsymbol{X}'\boldsymbol{X})^{-1})$]. 因而 A 最优设计最小化了回归系数的方差和.

因为许多响应曲面实验与响应的预测有关, 所以**预测方差准则**是实践中值得考虑的准则. **G 最优准则**是其中应用最广的一种. 一种设计称为 G 最优, 若它能最小化在整个设计区域内尺度化预测方差的最大值. 即,

$$\frac{N\mathrm{V}[\hat{y}(\boldsymbol{x})]}{\sigma^2}$$

在整个设计区域内的最大值最小, 其中 N 是设计中的点数. 如果模型有 p 个参数, 一个设计的 G 效率是

$$G_e = \frac{p}{\max \frac{N\mathrm{V}[\hat{y}(\boldsymbol{x})]}{\sigma^2}} \tag{11.22}$$

V 准则考虑设计区域内一个点集 (如 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m$) 的预测方差. 这个点集可能是所选设计的候选点集, 也可能是对实验者有特殊意义的点集. 使得在 m 个点上预测方差的均值最小的设计就是**V 最优设计**.

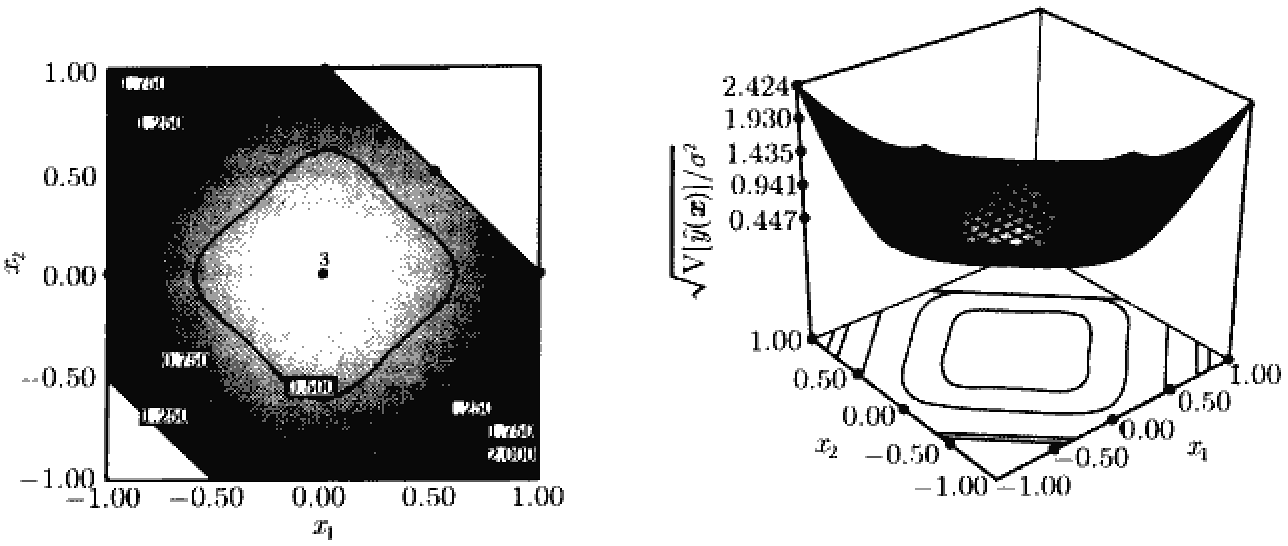
我们已讨论的设计准则统称为**字母最优准则**. 在有些情况下, 字母最优设计是已知的, 且可分析地构造. 2^k 设计就是一个这样的好例子, 它对于拟合有 k 个变量的一阶模型或有交互作用的一阶模型来讲, 是 D 最优设计、A 最优设计、G 最优设计及 V 最优设计. 然而, 在大多数情况下, 最优设计是未知的, 必须用计算机算法来寻找设计. 许多有实验设计模块的统计软件包都有这样的功能. 多数构造设计的程序用**交换算法**原理编程. 这种算法最简单的步骤是, 实验者先确定网格候选点, 并从中选择一个初始设计 (可以是随机的); 然后, 算法将在网格中但不在设计中的点与当前在设计中的点进行交换, 以改进所选的最优准则. 因为并不是每一个可能的设计都能被评估到, 所以不能保证一定可以找到最优设计, 然而, 交换算法程序一般可以得到一个 “接近” 最优结果的设计. 为了增加最后得到的设计非常接近最优设计的可能性, 有时需要重复进行几次从不同初始设计开始的设计构造过程.

为了说明这种想法, 考虑前面所讨论的如图 11.28 所示的不规则实验区域的粘合剂实验. 设分离力为响应变量, 想对此响应拟合二阶模型. 在图 11.29a 中显示了有 4 个中心点 (共 12 个试验) 内接该区域的中心复合设计. 它不是可旋转的设计, 但它是在此设计区域内可以拟合的最大 CCD. 此设计的 $|(\boldsymbol{X}'\boldsymbol{X})^{-1}| = 1.852\text{E-}2$, $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 的迹是 6.375. 图 11.29a 中还显示了预测响应标准差为常数的等高线图, 计算中假定 $\sigma = 1$. 图 11.29b 显示了相应的响应曲面图.

图 11.30a 和表 11.12 显示了此问题的有 12 个试验的 D 最优设计, 这个设计由 Design-Expert 软件包生成. 其 $|(\boldsymbol{X}'\boldsymbol{X})^{-1}| = 2.153\text{E-}4$. 在 D 准则下这个设计比内接 CCD 好得多. 内接 CCD 关于 D 最优设计的相对效率是

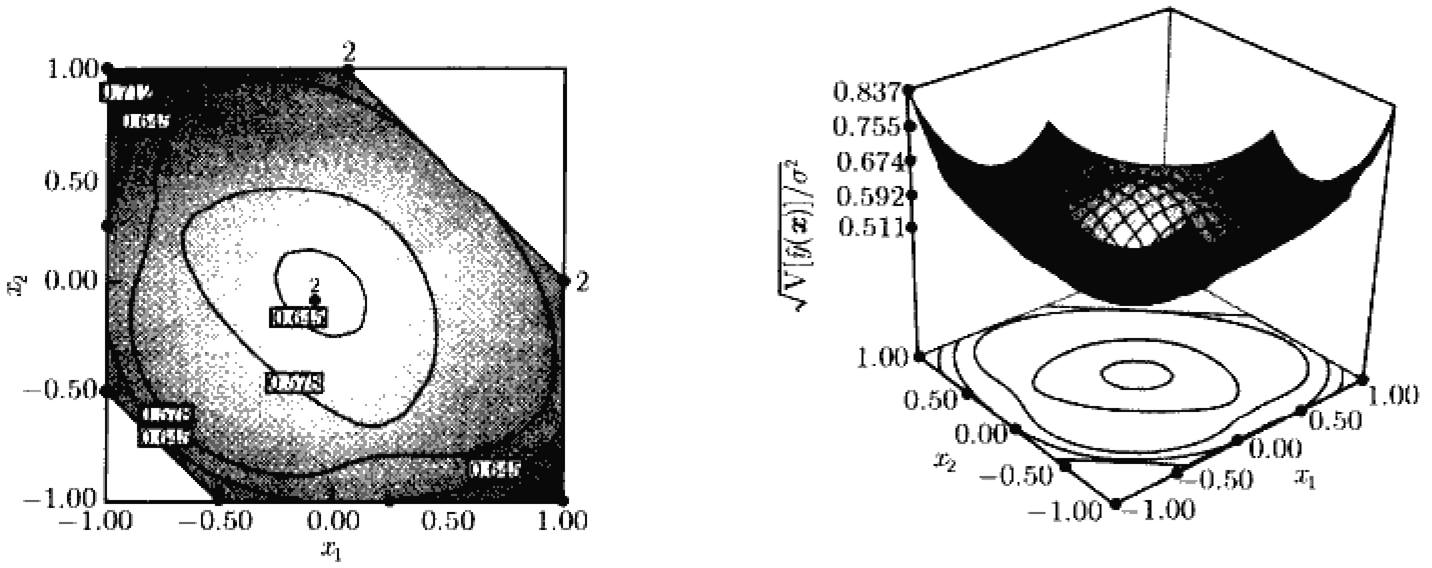
$$D_e = \left(\frac{|(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}|}{|(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}|}\right)^{1/p} = \left(\frac{0.000\ 215\ 3}{0.018\ 52}\right)^{1/6} = 0.476$$

也就是说, 内接 CCD 只有 D 最优设计 47.6%的效率. 这意味着 CCD 需要花 $1/0.476 = 2.1$ 倍 (近两倍) 的重复试验数以达到 D 最优设计所能达到的回归系数估计的精度. 这个 D 最优设计的 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 的迹是 2.516, 表明此设计的回归系数估计的方差之和比 CCD 小得多. 图 11.30a



(a) 设计与常数$\sqrt{V[\hat{y}(\boldsymbol{x})]/\sigma^2}$的等高线图　　(b) 响应曲面图

图 11.29　在图 11.28 的约束设计区域下的一个内接中心复合设计



(a) 设计与$\sqrt{V[\hat{y}(\boldsymbol{x})]/\sigma^2}$等高线　　(b) 响应曲面图

图 11.30　在图 11.28 的约束设计区域中的一个 D 最优设计

和图 11.30b 也显示了预测响应标准差为常数的等高线及相应的响应曲面图 (假定 $\sigma = 1$). D 最优设计预测标准差的等高线一般比内接 CCD 的要小, 特别是在靠近所关注区域的边界附近是如此, 因为内接 CCD 在那里没有设计点.

表 11.12　在图 11.26 的约束区域下的一个 D 最优设计

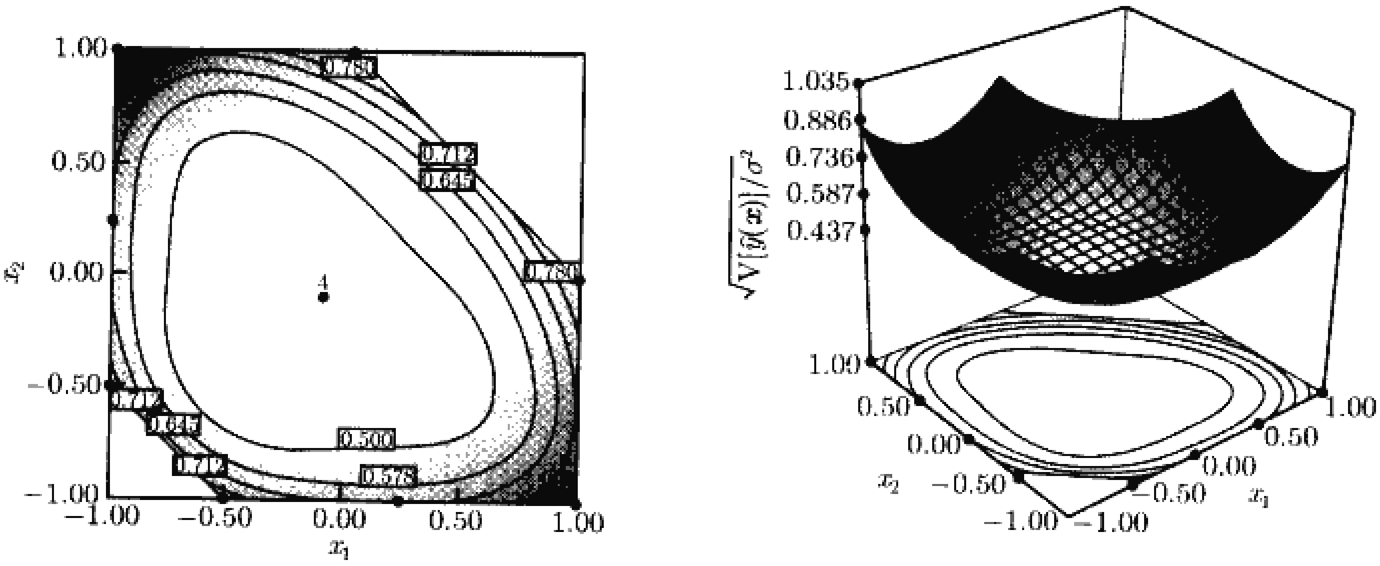
标准次序	x_1	x_2	标准次序	x_1	x_2
1	−0.50	−1.00	7	−1.00	0.25
2	1.00	0.00	8	0.25	−1.00
3	−0.08	−0.08	9	−1.00	−0.50
4	−1.00	1.00	10	1.00	0.00
5	1.00	−1.00	11	0.00	1.00
6	0.00	1.00	12	−0.08	−0.08

图 11.31a 显示了第 3 个设计, 此设计是将前面 D 最优设计的两个角上的重复点移到了中心点. 这可能是一个好主意, 因为图 11.30b 中显示的 D 最优设计预测响应的标准差在设计区

域的中间部分稍微增大了些. 图 11.31a 也显示了这个改进的 D 最优设计的预测标准差的等高线图, 图 11.31b 显示了响应曲面图. 这个设计的 D 准则是 $|(\boldsymbol{X}'\boldsymbol{X})^{-1}| = 3.71\text{E-}4$, 相对效率是

$$D_e = \left(\frac{|(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}|}{|(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}|}\right)^{1/p} = \left(\frac{0.000\ 215\ 3}{0.000\ 371}\right)^{1/6} = 0.91$$

即这个设计与 D 最优设计的效率几乎相同. $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 的迹是 2.473, 略小于 D 最优设计. 这个设计的预测标准差的等高线看起来至少像 D 最优设计一样好, 特别是在区域的中心.



(a) 设计与$\sqrt{V[\hat{y}(x)]/\sigma^2}$ 等高线　　(b) 响应曲面图

图 11.31　在图 11.28 的约束设计区域下的一个改进的 D 最优设计

计算机以字母最优准则产生的设计在实验区域既不是球体也不是立方体的情况下的确是有用的. 然而, 在多数问题中它们并不能取代标准设计. 因为在生成字母最优设计时, 只能严格遵守一个准则, 注意到在 11.4 节开始部分列出了几个不同的设计准则, 其中有几个准则实际上有点定性的或主观的. 在真正的实验问题中选择设计时, 通常需要评估许多准则. 有关这个问题更多的讨论, 见 Myers and Montgomery (2002, 第 8 章).

11.5　混料实验

前面各节所介绍的响应曲面设计中, 每个因子的水平都独立于另外因子的水平. 在**混料实验**中, 因子是混合物的分量或成分, 因此, 它们的水平不再是独立的. 例如, 如果 $x_1, x_2, \cdots, x_p$ 表示一混料实验的 p 个分量的比例, 则

$$0 \leqslant x_i \leqslant 1, \qquad i = 1, 2, \cdots, p$$

$$x_1 + x_2 + \cdots + x_p = 1(\text{即 } 100\%)$$

图 11.32 对 $p=2$ 和 $p=3$ 个分量的情形说明了这些约束. 对两个分量来说, 设计的因子空间包括两个分量在直线段 $x_1 + x_2 = 1$ 上的所有值, 每个分量界于 0 与 1 之间. 有 3 个分量时, 混料空间是一个三角形, 其顶点对应于**纯混合物** (仅由单一成分组成).

当混料实验有 3 个分量时, 为方便起见, 可以将约束的实验区域用**三线坐标纸**来表示, 如图 11.33 表示的那样. 图 11.33 的 3 条边中的每一条边表示 3 种成分中缺少一种成分 (此成分的名称标在此边对应的顶点处) 的混合物. 每一方向上的 9 条栅格线标记出分量的 10% 的增量.

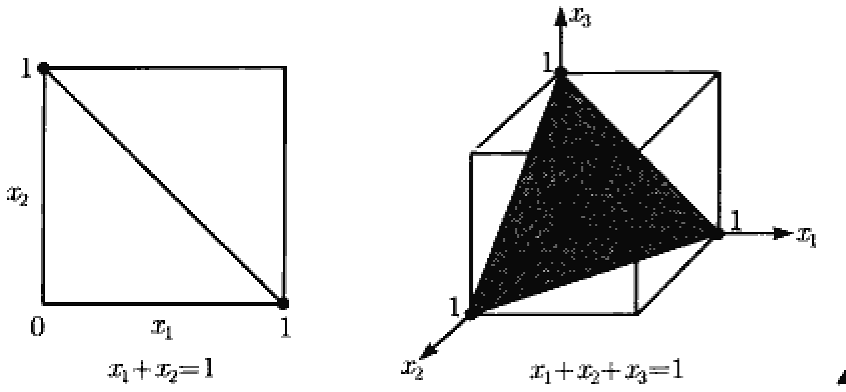


图 11.32 混料设计的约束因子空间

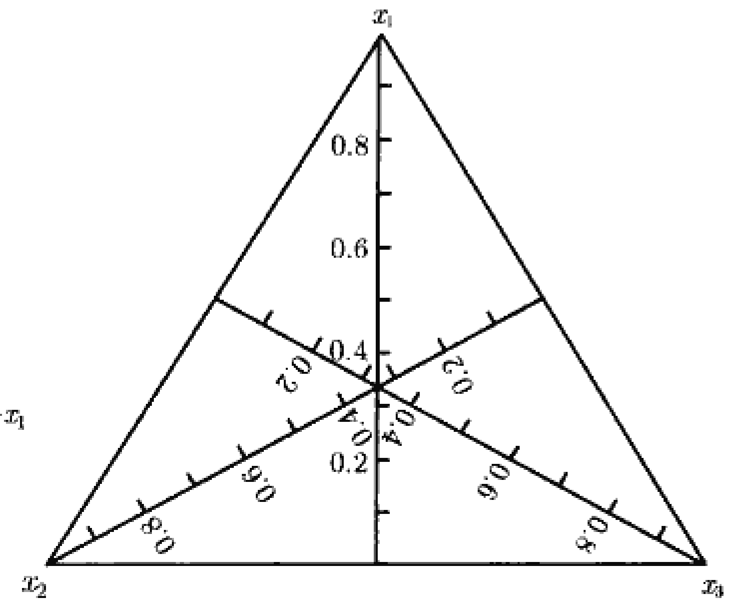


图 11.33 三线坐标系

单纯形设计用来研究混料分量在响应变量上的效应. p 个分量的$\{p, m\}$ **单纯形格点设计**由下述坐标值定义的点所组成: 每个分量的百分率取 0 到 1 之间的 $m+1$ 个等距离的值:

$$x_i = 0, \frac{1}{m}, \frac{2}{m}, \cdots, 1, \qquad i = 1, 2, \cdots, p \tag{11.23}$$

然后使用 (11.23) 式中比值的所有可能组合 (混合). 作为一个例子, 设 $p=3$ 与 $m=2$. 则

$$x_i = 0, \frac{1}{2}, 1, \qquad i = 1, 2, 3$$

而单纯形格点由下述 6 个试验所组成:

$$(x_1, x_2, x_3) = (1,0,0), (0,1,0), (0,0,1), \left(\frac{1}{2}, \frac{1}{2}, 0\right), \left(\frac{1}{2}, 0, \frac{1}{2}\right), \left(0, \frac{1}{2}, \frac{1}{2}\right)$$

此设计如图 11.34 所示, 3 个顶点 (1, 0, 0), (0, 1, 0), (0, 0, 1) 是纯混合物, 而 $(\frac{1}{2}, \frac{1}{2}, 0)$ $(\frac{1}{2}, 0, \frac{1}{2})$, $(0, \frac{1}{2}, \frac{1}{2})$ 是二元混合物或两种成分的混合物, 位于三角形 3 条边的中点上. 图 11.34 还显示出{3, 3}, {4, 2}, {4,3} 单纯形格点设计. 一般说来, $\{p, m\}$ 单纯形格点设计的点数是

$$N = \frac{(p+m-1)!}{m!(p-1)!}$$

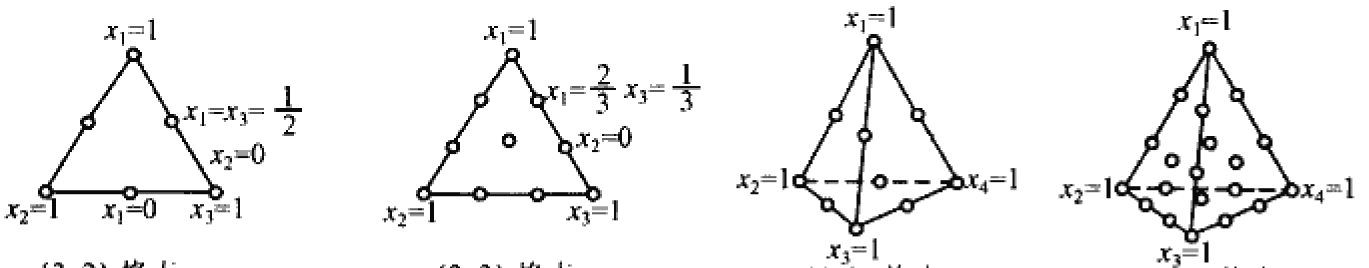


图 11.34 $p=3$ 和 $p=4$ 个分量的某些单纯形格点设计

另一种单纯形格点设计是**单纯形重心设计**. 在一个 p 分量单纯形重心设计中, 有 2^p-1 个点, 对应于 $(1, 0, \cdots, 0)$ 的 p 个排列, $(\frac{1}{2}, \frac{1}{2}, \cdots, 0)$ 的 $\binom{p}{2}$ 个排列, $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \cdots, 0)$ 的 $\binom{p}{3}$ 个排列, $\cdots$, 以及总重心 $(\frac{1}{p}, \frac{1}{p}, \cdots, \frac{1}{p})$. 图 11.35 显示了某些单纯形重心设计.

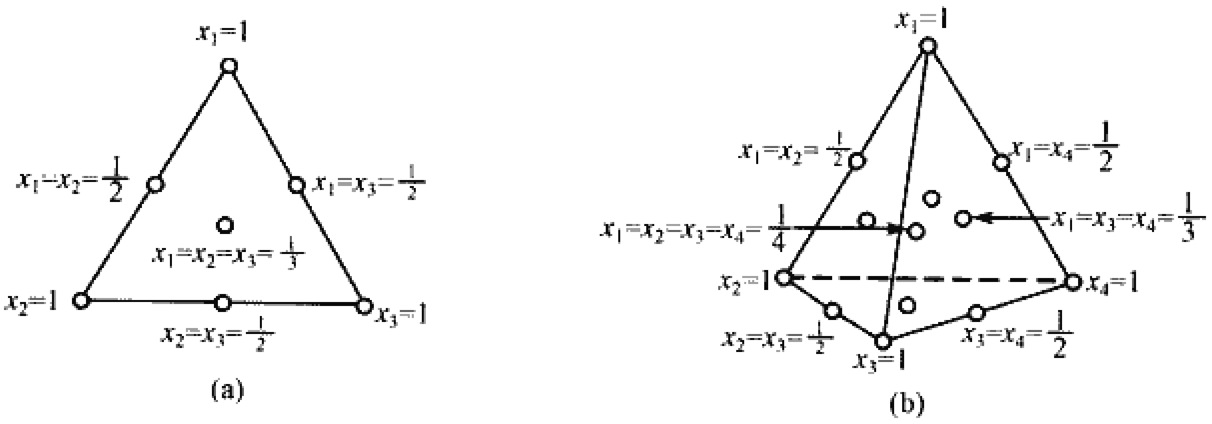


图 11.35 有 (a) $p=3$ 个分量 (b) $p=4$ 个分量的单纯形重心设计

对上面所描述的单纯形设计的一种批评是, 大多数试验点都是在区域的边界上出现, 因此只包含有 p 个成分中的 $p-1$ 个. 通常希望在单纯形格点设计或单纯形重心设计的区域内部加上一些附加点, 以使混合物能由所有 p 种成分组成. 进一步的讨论请参阅 Cornell (2002) 和 Myers and Montgomery (2002).

混料设计的模型不同于通常在响应曲面设计中所用的多项式, 因为有约束 $\sum x_i=1$. 广为使用的混料设计模型的标准形式有

线性的:

$$\mathrm{E}(y)=\sum_{i=1}^{p}\beta_i x_i \tag{11.24}$$

二次的:

$$\mathrm{E}(y)=\sum_{i=1}^{p}\beta_i x_i+\sum\sum_{i<j}^{p}\beta_{ij}x_i x_j \tag{11.25}$$

完全三次式:

$$\mathrm{E}(y)=\sum_{i=1}^{p}\beta_i x_i+\sum\sum_{i<j}^{p}\beta_{ij}x_i x_j+\sum\sum_{i<j}^{p}\delta_{ij}x_i x_j(x_i-x_j)+\sum\sum\sum_{i<j<k}\beta_{ijk}x_i x_j x_k \tag{11.26}$$

特殊三次式:

$$\mathrm{E}(y)=\sum_{i=1}^{p}\beta_i x_i+\sum\sum_{i<j}^{p}\beta_{ij}x_i x_j+\sum\sum\sum_{i<j<k}\beta_{ijk}x_i x_j x_k \tag{11.27}$$

这些模型的各项都有相对简单的解释. 从 (11.24) 式到 (11.27) 式, 参数 β_i 表示纯混合物 $x_i=1$ 和 $x_j=0$ 当 $j\neq i$ 时的期望响应. $\sum_{i=1}^{p}\beta_i x_i$ 部分称为**线性混合物**部分. 当分量对之间的非线性混合物引起弯曲时, 参数 β_{ij} 或者表示**协同性的混合**或者表示**对立性的混合**. 在混料模型中经常需要更高阶的项, 因为 (1) 所研究的现象可能很复杂, (2) 实验的区域经常是整个可操作的区域, 因而是较大的区域, 需要有一个更精致的模型.

例 11.3 有 3 个分量的混料设计

Cornell (2002) 描述了一个混料实验. 其中混合了 3 种成分 —— 聚乙烯 (x_1)、聚苯乙烯 (x_2) 和聚丙烯 (x_3)—— 并将其制成纤维, 用以纺成织布用的纱. 感兴趣的响应变量是在数千克的力作用下纱的伸长值. 用一个{3, 2} 单纯形格点设计来研究产品. 设计和观测得到的响应如表 11.13 所示. 注意, 所有的格点或者是纯混合物, 或者是二元混合物, 也就是说, 产品的任一配方仅用了 3 种成分中的最多两种. 进行重复

观测，每种纯混合物进行两次重复，每种二元混合物进行 3 次重复. 误差的标准差可以用这些重复观测来估计，$\hat{\sigma} = 0.85$. Cornell 用二阶混合多项式来拟合这些数据，得到

$$\hat{y} = 11.7x_1 + 9.4x_2 + 16.4x_3 + 19.0x_1x_2 + 11.4x_1x_3 - 9.6x_2x_3$$

这能够说明此模型是响应的合适表示式. 因为 $\hat{\beta}_3 > \hat{\beta}_1 > \hat{\beta}_2$，结论是，成分 3 (聚丙烯) 生产的纱有最高的伸长值. 而且，因为 $\hat{\beta}_{12}$ 和 $\hat{\beta}_{13}$ 是正的，成分 1 和 2 或成分 1 和 3 的混合物产生较高的伸长值，比纯混合物的平均伸长值要高. 这是一个"协同性"混合效应的例子. 因为 $\hat{\beta}_{23}$ 是负的，所以成分 2 和 3 有对立性的混合效应.

表 11.13 纱伸张度问题的{3, 2} 单纯形格点设计

设计点	成分比例			观测到的伸长值	平均伸长值
	x_1	x_2	x_3		$\bar{y}$
1	1	0	0	11.0, 12.4	11.7
2	$\frac{1}{2}$	$\frac{1}{2}$	0	15.0, 14.8, 16.1	15.3
3	0	1	0	8.8, 10.0	9.4
4	0	$\frac{1}{2}$	$\frac{1}{2}$	10.0, 9.7, 11.8	10.5
5	0	0	1	16.8, 16.0	16.4
6	$\frac{1}{2}$	0	$\frac{1}{2}$	17.7, 16.4, 16.6	16.9

图 11.36 画出了伸长值的等高线图，这有助于解释这些结果. 审视一下这张图，如果希望伸长值达到最大，则应选择成分 1 和 3 的混合物，而且由约 80%的成分 3 和 20%的成分 1 所构成.

前面已注意到单纯形格点设计和单纯形重心设计都是**边界点设计**. 如果实验者想要作出完全混料性质的预测，则在单纯形内部做更多的试验是很理想的. 我们推荐在普通单纯形设计中增加**轴试验**和总重心处的试验 (如果总重心还不是设计点).

分量 i 的轴是从基点 $x_i = 0$, $x_j = 1/(p-1)$(对一切 $j \neq i$) 出发，到对顶点 $x_i = 1$, $x_j = 0$ (对一切 $j \neq i$) 的一条直线或射线. 基点总是位于单纯形的顶点 $x_i = 1$, $x_j = 0$(对一切 $j \neq i$) 的对面的 $p-2$ 维边缘的重心处. [这个边缘有时也称为 $(p-2)$ 维平面.] 分量轴的长度是一个单位. **轴点**放在分量轴上距离重心 Δ 处. Δ 的最大值是 $(p-1)/p$. 建议将轴试验点放置在单纯形的重心与每个顶点的中间，所以 $\Delta = (p-1)/2p$. 有时称这些点为**轴检测混合**，因为，在实践中通常在拟合预备混料模型并用这些轴点处的响应来检测预备模型的合适性之后，就将这些点排除.

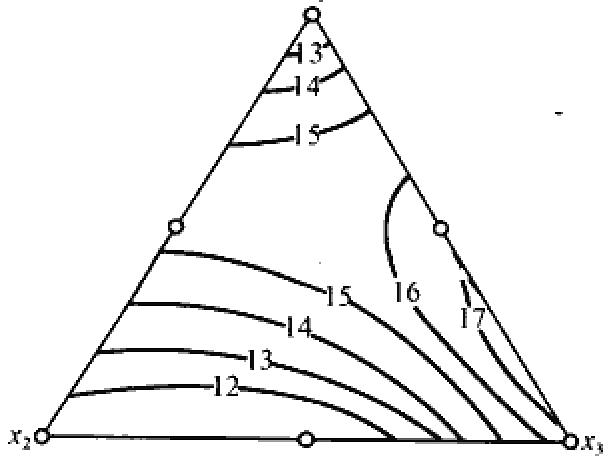


图 11.36 例 11.3 二阶混料模型所估计的纱的伸长值的等高线图

图 11.37 显示了增加了轴点的{3, 2} 单纯形格点设计. 这个设计有 10 个点, 其中有 4 个点在单纯形内部. {3, 3} 单纯形格点设计可支持拟合完全三次模型, 但扩大的单纯形格点设计并不如此; 然而扩大的单纯形格点可以让实验者拟合特殊三次模型, 或者在二次模型中添加诸如 $\beta_{1233}x_1x_2x_3^2$ 的特别的 4 次项. 从探测和建立完全三次模型中不能计算的三角形内部的弯曲性的意义来说, 扩大的单纯形格点设计对于研究完全混料的响应有良好的表现. 扩大的单纯形格点设计在检测拟合不足时比{3, 3} 格点设计更强. 这类设计在下列情况下特别有用, 即, 实验者不能确定什么样的模型更合适, 需要序贯地建立模型, 也就是要从简单的多项式模型 (或许是一阶模型) 开始, 检验模型的拟合不足, 然后用更高阶的项扩大模型, 检验新模型的拟合不足, 等等.

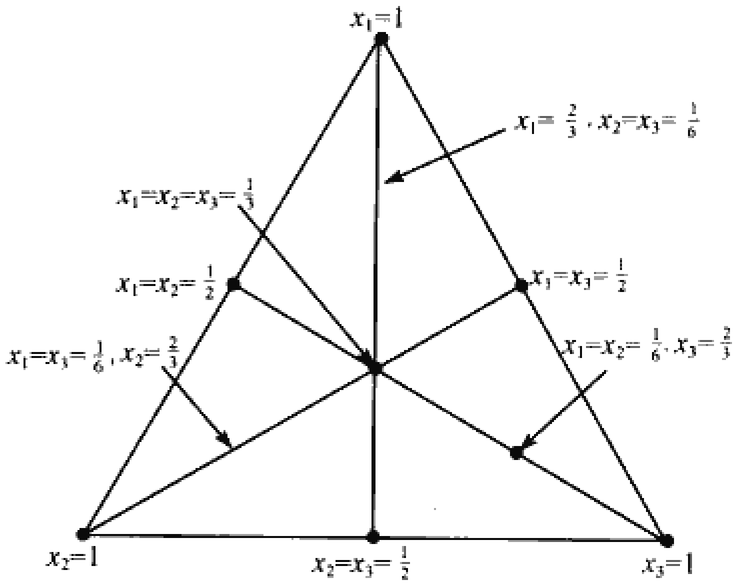


图 11.37　一个扩大的单纯形格点设计

在某些混料问题中, 会出现对各个分量的**约束**. 形如

$$l_i \leqslant x_i \leqslant 1, \qquad i = 1, 2, \cdots, p$$

的下界约束十分普遍. 在只有下界约束时, 可行设计区域仍然是单纯形, 但它内接于原单纯形区域内. 这种情况可以通过引用定义为

$$x_i' = \frac{x_i - l_i}{\left(1 - \sum_{j=1}^{p} l_j\right)} \tag{11.28}$$

的**伪分量** (pseudocomponent) 来简化, 其中 $\sum_{j=1}^{p} l_j < 1$. 现有

$$x_1' + x_2' + \cdots + x_p' = 1$$

所以, 当下界是实验情况的一部分时, 可通过引进伪分量来使用单纯形设计. 利用 (11.28) 式的逆变换式, 就可以将伪分量的单纯形设计的特定配方变换为原来分量的配方. 也就是说, 如果 x_i' 是一次试验中第 i 个伪分量的设计值, 则第 i 个原来的混合分量是

$$x_i = l_i + \left(1 - \sum_{j=1}^{p} l_j\right) x_i' \tag{11.29}$$

如果一个分量同时有上界约束也有下界约束, 则可行区域不再是单纯形, 而是一个不规则多边形. 因为实验区域不再是标准的形状, 所以**计算机生成设计**对这类混料问题很有用.

例 11.4 涂料配方

实验者想优化汽车透明面漆的配方. 面漆是合成产品, 且有特殊的性能要求. 特别地, 顾客需要 Knoop 硬度超过 25、固体物质比例低于 30%的面漆. 透明面漆由 3 种成分混合组成, 它们是单体 (x_1)、交联剂 (x_2) 和树脂 (x_3). 对成分比例有如下约束:

$$x_1 + x_2 + x_3 = 100$$
$$5 \leqslant x_1 \leqslant 25$$
$$25 \leqslant x_2 \leqslant 40$$
$$50 \leqslant x_3 \leqslant 70$$

约束的实验区域如图 11.38 所示. 因为这个区域并不是单纯形, 我们用 D 最优设计来求解这个问题. 假定两个响应都用二次混料模型来建模, 我们可以用 Design-Expert 得到如图 11.38 所示的 D 最优设计. 假定除了附加 6 个二次混料模型所需的试验外, 另做 4 个不同的试验以检测拟合不足, 这些试验中有 4 个试验要做重复试验以提供纯误差的估计. Design-Expert 用顶点、棱的中点、区域重心以及检测试验点 (位于重心与顶点中间) 作为候选点.

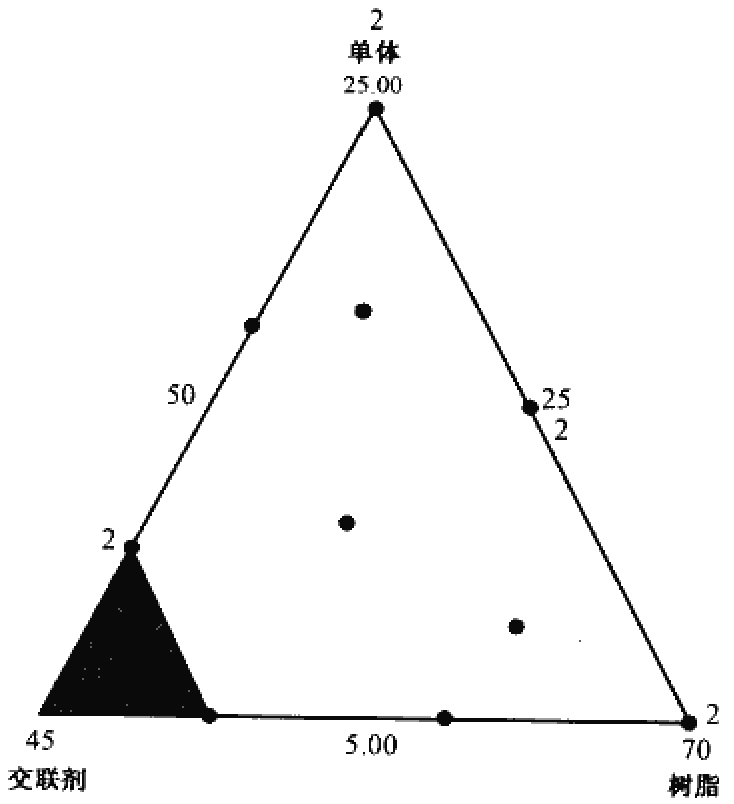


图 11.38 例 11.4 面漆配方问题的约束实验区域 (以实际分量的尺度显示)

含有 14 个试验的设计, 以及硬度和固体物质响应都列在表 11.14 中. 对两个响应拟合的二次模型的结果小结在表 11.15 和表 11.16 中. 注意到, 对于硬度与固体物质响应, 二次模型都拟合得很好. 两个响应 (关于伪分量) 的拟合方程也都列在表中. 图 11.39 和图 11.40 显示了响应的等高线图.

表 11.14 例 11.4 面漆配方问题的 D 最优设计

标准序号	试验序号	单体 x_1	交联剂 x_2	树脂 x_3	硬度 y_1	固体物质 y_2
1	2	17.50	32.50	50.00	29	9.539
2	1	10.00	40.00	50.00	26	27.33
3	4	15.00	25.00	60.00	17	29.21
4	13	25.00	25.00	50.00	28	30.46
5	7	5.00	25.00	70.00	35	74.98
6	3	5.00	32.50	62.50	31	31.5
7	6	11.25	32.50	56.25	21	15.59

(续)

标准序号	试验序号	单体 x_1	交联剂 x_2	树脂 x_3	硬度 y_1	固体物质 y_2
8	11	5.00	40.00	55.00	20	19.2
9	10	18.13	28.75	53.13	29	23.44
10	14	8.13	28.75	63.13	25	32.49
11	12	25.00	25.00	50.00	19	23.01
12	9	15.00	25.00	60.00	14	41.46
13	5	10.00	40.00	50.00	30	32.98
14	8	5.00	25.00	70.00	23	70.95

表 11.15 关于硬度响应的模型拟合

```
Response:  hardness
  ANOVA for Mixture Quadratic Model
Analysis of variance table [Partial sum of squares]
                            Sum of                   Mean         F
       Source               Squares     DF          Square      Value      Prob>F
Model                       279.73       5           55.95       2.37      0.1329
Linear Mixture               29.13       2           14.56       0.62      0.5630
   AB                        72.61       1           72.61       3.08      0.1174
   AC                       179.67       1          179.67       7.62      0.0247
   BC                         8.26       1            8.26       0.35      0.5703
Residual                    188.63       8           23.58
  Lack of Fit                63.63       4           15.91       0.51      0.7354
  Pure Error                125.00       4           31.25
Cor Total                   468.36      13

Std.Dev.                      4.86         R-Squared                 0.5973
Mean                         24.79         Adj R-Squared             0.3455
C.V.                         19.59         Pred R-Squared           -0.3635
PRESS                       638.60         Adeq Precision             4.975

                       Coefficient                Standard     95% Cl      95% Cl
Component                Estimate       DF          Error        Low        High
A-Monomer                  23.81         1           3.36       16.07       31.55
B-Crosslinker              16.40         1           7.68       -1.32       34.12
C-Resin                    29.45         1           3.36       21.71       37.19
AB                         44.42         1          25.31      -13.95      102.80
AC                        -44.01         1          15.94      -80.78       -7.25
BC                         13.80         1          23.32      -39.97       67.57
Final Equation in Terms of Pseudo Components:
                          hardness=
                           +23.81*A
                           +16.40*B
                           +29.45*C
                           +44.42*A*B
                           -44.01*A*C
                           +13.80*B*C
```

表 11.16 关于固体物质响应的模型拟合

Response: solids
ANOVA for Mixture Quadratic Model
Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob>F
Model	4297.92	5	859.59	25.78	<0.0001
Linear Mixture	*2931.09*	*2*	*1465.66*	*43.95*	*<0.0001*
AB	*211.20*	*1*	*211.20*	*6.33*	*0.0360*
AC	*285.67*	*1*	*285.67*	*8.57*	*0.0191*
BC	*1036.72*	*1*	*1036.72*	*31.09*	*0.0005*
Residual	266.79	8	33.35		
Lack of Fit	*139.92*	*4*	*34.98*	*1.10*	*0.4633*
Pure Error	*126.86*	*4*	*31.72*		
Cor Total	4564.73	13			

Std. Dev.	5.77	R-Squared	0.9416
Mean	33.01	Adj R-Squared	0.9050
C.V.	17.49	Pred R-Squared	0.7827
PRESS	991.86	Adeq Precision	15.075

Component	Coefficient Estimate	DF	Standard Error	95% Cl Low	95% Cl High
A-Monomer	26.53	1	3.99	17.32	35.74
B-Crosslinker	46.60	1	9.14	25.53	67.68
C-Resin	73.23	1	3.99	64.02	82.43
AB	-75.76	1	30.11	-145.19	-6.34
AC	-55.50	1	18.96	-99.22	-11.77
BC	-154.61	1	27.73	-218.56	-90.67

Final Equation in Terms of Pseudo Components:

solids=
+26.53*A
+46.60*B
+73.23*C
-75.76*A*B
-55.50*A*C
-154.61*B*C

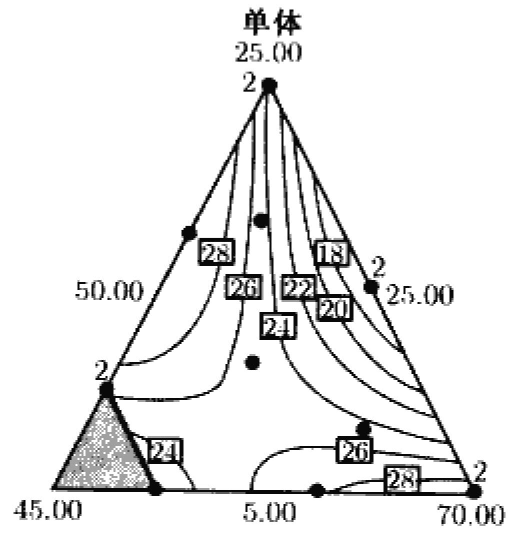


图 11.39 例 11.4 中 Knoop 硬度响应的等高线图

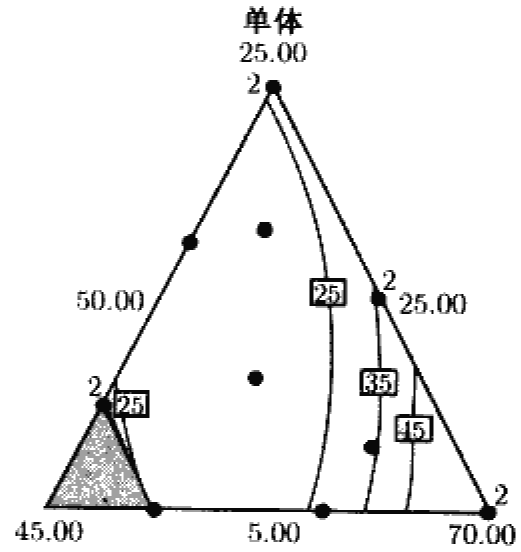


图 11.40 例 11.4 中固体物质比例响应的等高线图

图 11.41 是两个响应曲面的重叠等高线图, 显示了 Knoop 硬度的 25 等高线和固体物质 30%等高线. 产品的可行域是图的中心处没有阴影部分. 显然, 在透明面漆能满足性能要求的条件下, 单体、交联剂及树脂的比例有相当大的选择余地.

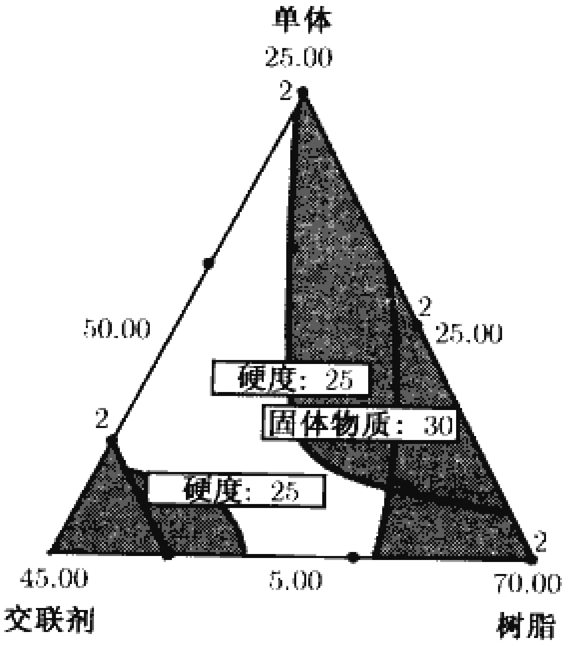


图 11.41 Knoop 硬度与固体物质比例响应在面漆配方的可行域内的重叠图

11.6 调优运算

响应曲面法常在小型试验工厂中使用, 用以研究和开发. 当应用到大规模的生产过程中时, 因为实验程序相对比较复杂, 所以通常只做一次 (或很不经常做). 但是, 在小型试验工厂中得到的最优条件不一定是大规模生产时的最优条件. 小型工厂也许每天生产 2 磅产品, 而大规模生产过程每天也许就要生产 2 000 磅. 小型试验工厂的 "规模上升" 至大规模生产过程, 通常会导致最优条件的变化. 甚至当大规模工厂以此最优条件运行时, 因为原料的变异、环境的改变、操作人员的改变等原因, 最后也会从那一最优条件 "漂移" 开来.

我们需要有一种方法来继续监视并改善大规模生产过程. 其目标是将运行条件自动转为最优条件或者自动调节 "漂移". 这种方法不应该对运行条件作大的更改或突然更改, 那样做可能会使生产停顿. Box (1957) 提出的**调优运算 (EVOP)** 就是这类操作方法. 它是作为常规性工厂运行方法来设计的, 在研究和开发人员稍加帮助下, 就可以由生产人员来执行.

EVOP 由对所考虑的运行变量水平有计划地引进小的改变所组成. 通常, 用 2^k 设计来做到这一点. 取变量的小改变量不会对产量、质量或数量产生大的干扰, 而变量的大改变量会使过程性能可能的改善最终被发现. 对所感兴趣的响应变量在 2^k 设计的每个点处收集数据. 当在每个设计点处都取得一个观测值时, 就叫做完成了一个**循环**. 然后计算过程变量的效应和交互作用. 最后, 经过几个循环之后, 一个或多个过程变量的效应或它们的交互作用就可能显示出对响应有显著的效应. 此时, 就应该作出决定, 来改变基本运行条件以改善响应. 当改进的条件被检测而通过时, 就称为完成了一个**阶段**.

在检测过程变量和交互作用的显著性时, 需要实验误差的估计量. 这可从循环数据中算得. 此外, 2^k 设计通常以当前最好的运行条件为中心. 将这一点上的响应和析因设计部分的 2^k 个点上的响应进行比较, 就可以检测弯曲性或**平均改变量** (CIM); 也就是说, 如果过程实际上以最大值为中心, 则在中心点处的响应应该显著地大于在 2^k 个周边点处的响应.

从理论上说, EVOP 可应用于 k 个过程变量. 在实践中, 通常只考虑两个或 3 个变量. 我们对此方法将提供一个有两个变量的例子. Box and Draper (1969) 对 3 个变量的情况进行了详尽的讨论, 包括必要的形式和工作纸. Myers and Montgomery (2002) 讨论了 EVOP 的计算机实现方法.

例 11.5 考虑一化学过程, 其产率是温度 (x_1) 和压强 (x_2) 的函数. 当前的操作条件是 $x_1 = 250$ °F 和 $x_2 = 145$ psi. EVOP 方法用 2^2 设计加中心点, 如图 11.42 所示. 在每个设计点上按数字顺序 (1, 2, 3, 4, 5) 做完试验就完成了一个循环. 第 1 个循环的产率如图 11.42 所示.

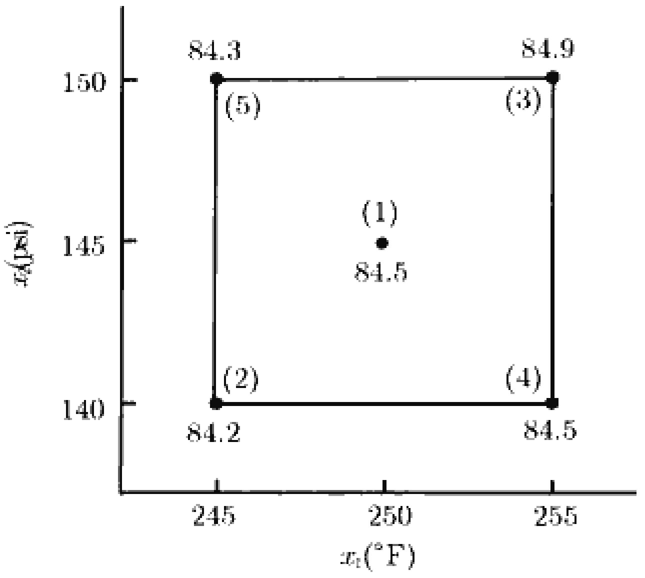
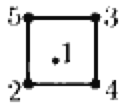


图 11.42 EVOP 的 2^2 设计

将第 1 个循环所得的产率记录在 EVOP 计算纸上, 如表 11.17 所示. 在第 1 个循环的计算纸的末尾, 作不出标准差的估计量. 温度和压强的效应和交互作用按 2^2 设计的方法计算.

表 11.17 例 11.5 的 EVOP 计算纸 ($n = 1$)



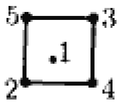
	循环: $n = 1$ 响应: 产率					阶段: 1 日期: 1/11/04
	计算平均值					计算标准差
运行条件	(1)	(2)	(3)	(4)	(5)	
(i) 前一循环的和						前一和 $S =$
(ii) 前一循环的平均值						前一平均值 $S =$
(iii) 新的观测值	84.5	84.2	84.9	84.5	84.3	新的 $S =$ 极差 $\times f_{5,n} =$
(iv) 差 [(ii) − (iii)]						(iv) 的极差 $=$
(v) 新的和 [(i) + (iii)]	84.5	84.2	84.9	84.5	84.3	新的和 $S =$
(vi) 新的平均值 [$\bar{y}_i =$ (v) / n]	84.5	84.2	84.9	84.5	84.3	新的平均值 $S = \frac{\text{新的和}S}{n-1}$

计算效应	计算误差限
温度效应 $= \frac{1}{2}(\bar{y}_3 + \bar{y}_4 - \bar{y}_2 - \bar{y}_5) = 0.45$	对新的平均值 $= \frac{2}{\sqrt{n}}S =$
压强效应 $= \frac{1}{2}(\bar{y}_3 + \bar{y}_5 - \bar{y}_2 - \bar{y}_4) = 0.25$	对新的效应 $\frac{2}{\sqrt{n}}S =$
$T \times P$ 交互效应 $= \frac{1}{2}(\bar{y}_2 + \bar{y}_3 - \bar{y}_4 - \bar{y}_5) = 0.15$	对平均改变量 $\frac{1.78}{\sqrt{n}}S =$
平均改变量的效应 $= \frac{1}{5}(\bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_5 - 4\bar{y}_1) = 0.02$	

然后进行第 2 个循环的试验, 并在另一张 EVOP 的计算纸上记录产率的数据, 如表 11.18 所示. 在第

2 个循环的计算纸末尾, 可以估计实验的误差, 并将效应的估计量与近似的 95% (两个标准差) 误差限进行比较. 极差是行 (iv) 的差的极差, 所以极差是 $+1.0-(-1.0)=2.0$. 因为表 11.18 的效应没有超过它们的误差限, 所以真实的效应可能是零, 从而不改变运行条件.

表 11.18　例 11.5 的 EVOP 计算纸 ($n=2$)

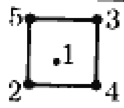
	循环: $n=2$ 响应: 产率	阶段: 1 日期: 1/11/04

	计算平均值					计算标准差
运行条件	(1)	(2)	(3)	(4)	(5)	
(i) 前一循环的和	84.5	84.2	84.9	84.5	84.3	前一和 $S=$
(ii) 前一循环的平均值	84.5	84.2	84.9	84.5	84.3	前一平均值 $S=$
(iii) 新的观测值	84.9	84.6	85.9	83.5	84.0	新的 $S=$ 极差 $\times f_{5,n}=0.60$
(iv) 差 [(ii) − (iii)]	−0.4	−0.4	−1.0	+1.0	0.3	(iv) 的极差 $=2.0$
(v) 新的和 [(i) + (iii)]	169.4	168.8	170.8	168.0	168.3	新的和 $S=0.60$
(vi) 新的平均值 [$\bar{y}_i=$ (v) / n]	84.70	84.40	85.40	84.00	84.15	新的平均值 $S=\frac{\text{新的和}S}{n-1}=0.60$

计算效应	计算误差限
温度效应 $=\frac{1}{2}(\bar{y}_3+\bar{y}_4-\bar{y}_2-\bar{y}_5)=0.43$	对新的平均值 $=\frac{2}{\sqrt{n}}S=0.85$
压强效应 $=\frac{1}{2}(\bar{y}_3+\bar{y}_5-\bar{y}_2-\bar{y}_4)=0.58$	对新的效应 $\frac{2}{\sqrt{n}}S=0.85$
$T\times P$ 交互效应 $=\frac{1}{2}(\bar{y}_2+\bar{y}_3-\bar{y}_4-\bar{y}_5)=0.83$	
平均改变量的效应 $=\frac{1}{5}(\bar{y}_2+\bar{y}_3+\bar{y}_4+\bar{y}_5-4\bar{y}_1)=-0.17$	对平均改变量 $\frac{1.78}{\sqrt{n}}S=0.76$

第 3 个循环的结果如表 11.19 所示. 现在压强的效应超过它的误差限, 温度效应等于误差限. 现在, 看来应该改变运行条件了.

表 11.19　例 11.5 的 EVOP 计算纸 ($n=3$)

	循环: $n=3$ 响应: 产率	阶段: 1 日期: 1/11/04

	计算平均值					计算标准差
运行条件	(1)	(2)	(3)	(4)	(5)	
(i) 前一循环的和	169.4	168.8	170.8	168.0	168.3	前一和 $S=0.60$
(ii) 前一循环的平均值	84.70	84.40	85.40	84.00	84.15	前一平均值 $S=0.60$
(iii) 新的观测值	85.0	84.0	86.6	84.9	85.2	新的 $S=$ 极差 $\times f_{5,n}=0.56$
(iv) 差 [(ii) − (iii)]	−0.30	+0.40	−1.20	−0.90	−1.05	(iv) 的极差 $=1.60$
(v) 新的和 [(i) + (iii)]	254.4	252.8	257.4	252.9	253.5	新的和 $S=1.16$
(vi) 新的平均值 [$\bar{y}_i=$(v)/n]	84.80	84.27	85.80	84.30	84.50	新的平均值 $S=\frac{\text{新的和}S}{n-1}=0.58$

计算效应	计算误差限
温度效应 $=\frac{1}{2}(\bar{y}_3+\bar{y}_4-\bar{y}_2-\bar{y}_5)=0.67$	对新的平均值 $\frac{2}{\sqrt{n}}S=0.67$
压强效应 $=\frac{1}{2}(\bar{y}_3+\bar{y}_5-\bar{y}_2-\bar{y}_4)=0.87$	对新的效应 $\frac{2}{\sqrt{n}}S=0.67$
$T\times P$ 交互效应 $=\frac{1}{2}(\bar{y}_2+\bar{y}_3-\bar{y}_4-\bar{y}_5)=0.64$	
平均改变量的效应 $=\frac{1}{5}(\bar{y}_2+\bar{y}_3+\bar{y}_4+\bar{y}_5-4\bar{y}_1)=-0.07$	对平均改变量 $\frac{1.78}{\sqrt{n}}S=0.60$

看一下这些结果, 似乎有理由关于点 (3) 开始新的 EVOP 阶段. 于是, $x_1=255$ °F, $x_2=150$ psi 就

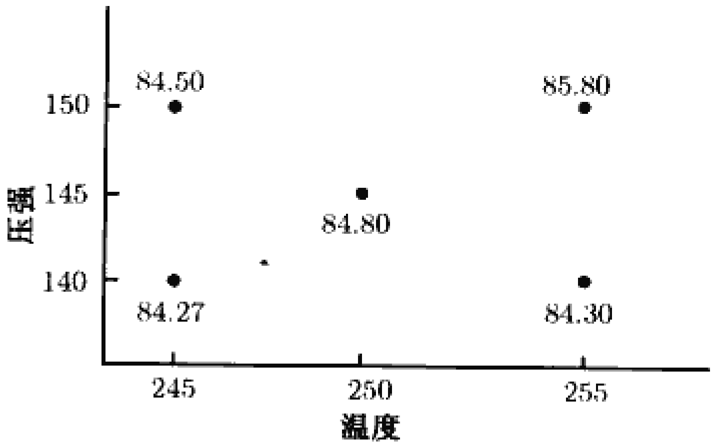
成为 2^2 设计在第 2 个阶段中的中心点.

EVOP 的一个重要方面就是将产生的信息反馈给过程操作人员和管理者. 这一点可通过安放在醒目位置上的 EVOP 信息板来完成. 本例在第 3 个循环之后的信息板如表 11.20 所示.

表 11.20 EVOP 信息板 (循环 3)

响应: 产率

要求: 最大化



	平均值的误差限:	±0.67	
效应及其 95%误差限	温度	0.67	±0.67
	压强	0.87	±0.67
	$T \times P$	0.64	±0.67
	平均改变量	0.07	±0.60
标准差	0.58		

EVOP 计算纸上大多数的量直接由分析 2^2 析因设计而得到. 例如, 任一效应的方差, 比如, $\frac{1}{2}(\bar{y}_3+\bar{y}_5-\bar{y}_2-\bar{y}_4)$ 的方差, 简化是 σ^2/n, 其中 σ^2 是观测值 (y) 的方差. 这样一来, 任一效应上的两个标准差误差限 (对应于 95%) 将是 $\pm 2\sigma/\sqrt{n}$. 平均改变量的方差是

$$\mathrm{V(CIM)} = \mathrm{V}[\frac{1}{5}(\bar{y}_2+\bar{y}_3+\bar{y}_4+\bar{y}_5-4\bar{y}_1)] = \frac{1}{25}(4\sigma_{\bar{y}}^2+16\sigma_{\bar{y}}^2) = \left(\frac{20}{25}\right)\frac{\sigma^2}{n}$$

于是, CIM 上的两个标准差误差限是 $\pm(2\sqrt{20/25})\sigma/\sqrt{n} = \pm 1.78\sigma/\sqrt{n}$.

标准差 σ 用极差法估计. 令 $y_i(n)$ 表示循环 n 的第 i 个设计点处的观测值, $\bar{y}_i(n)$ 表示经 n 个循环之后的对应于 $y_i(n)$ 的平均值. EVOP 计算纸中行 (iv) 的量是差 $y_i(n)-\bar{y}_i(n-1)$. 这些差的方差是

$$\mathrm{V}[y_i(n)-\bar{y}_i(n-1)] \equiv \sigma_D^2 = \sigma^2\left[1+\frac{1}{(n-1)}\right] = \sigma^2\frac{n}{(n-1)}$$

差的极差 (记作 R_D) 与差的标准差的估计量的关系是 $\hat{\sigma}_D = R_D/d_2$. 因子 d_2 依赖于用来计算 R_D 的观测值的个数. 现在 $R_D/d_2 = \hat{\sigma}\sqrt{n/(n-1)}$, 所以

$$\hat{\sigma} = \sqrt{\frac{(n-1)}{n}}\frac{R_D}{d_2} = (f_{k,n})R_D \equiv S$$

可用于估计观测值的标准差, 其中 k 表示用于设计的点数. 对于有一个中心点的 2^2 设计, $k=5$; 有一个中心点的 2^3 设计, $k=9$. $f_{k,n}$ 的值见表 11.21.

表 11.21　$f_{k,n}$ 的值

$n=$	2	3	4	5	6	7	8	9	10
$k=5$	0.30	0.35	0.37	0.38	0.39	0.40	0.40	0.40	0.41
9	0.24	0.27	0.29	0.30	0.31	0.31	0.31	0.32	0.32
10	0.23	0.26	0.28	0.29	0.30	0.30	0.30	0.31	0.31

11.7　思　考　题

*11.1　一化学工厂利用先液化空气, 再用分馏法分解的方法生产氧气. 氧气的纯度是主冷凝器温度和分馏塔上下之间压强比的函数. 当前的操作条件是温度 $(\xi_1)=-220$℃, 压强比 $(\xi_2)=1.2$. 利用下列数据, 求最速上升路径.

温度 (ξ_1)	−225	−225	−215	−215	−220	−220	−220	−220
压强比 (ξ_2)	1.1	1.3	1.1	1.3	1.2	1.2	1.2	1.2
纯度	82.8	83.5	84.7	85.0	84.1	84.5	83.9	84.3

11.2　一位工业工程师开发双项存货系统的计算机模拟模型. 决策变量是每一项目的订货量和再订购点. 需要最小化的响应是总存货的成本. 模拟模型用来产生如下表所示的数据. 识别出实验的设计. 求最速上升路径.

项目 1		项目 2		
订货量 (ξ_1)	再订购点 (ξ_2)	订货量 (ξ_3)	再订购点 (ξ_4)	总成本
100	25	250	40	625
140	45	250	40	670
140	25	300	40	663
140	25	250	80	654
100	45	300	40	648
100	45	250	80	634
100	25	300	80	692
140	45	300	80	686
120	35	275	60	680
120	35	275	60	674
120	35	275	60	681

11.3　证明下列设计是一单纯形设计. 拟合一阶模型并求最速上升路径.

x_1	x_2	x_3	y
0	$\sqrt{2}$	−1	18.5
$-\sqrt{2}$	0	1	19.8
0	$-\sqrt{2}$	−1	17.4
$\sqrt{2}$	0	1	22.5

11.4　求一阶模型

$$\hat{y}=60+1.5x_1-0.8x_2+2.0x_3$$

的最速上升路径. 变量规范为 $-1\leqslant x_i\leqslant 1$.

11.5 3 个因子的实验区域是时间 ($40 \leqslant T_1 \leqslant 80$ min)、温度 ($200 \leqslant T_2 \leqslant 300$℃) 和压强 ($20 \leqslant P \leqslant 50$ psig). 规范变量的一阶模型已很好地拟合了 2^3 设计的产率数据. 所得模型是

$$\hat{y} = 30 + 5x_1 + 2.5x_2 + 3.5x_3$$

点 $T_1 = 85$, $T_2 = 325$, $P = 60$ 是否在最速上升路径上?

11.6 两个因子的实验区域是温度 ($100 \leqslant T \leqslant 300$ ℉) 和催化剂进料速率 ($10 \leqslant C \leqslant 30$ lb/in). 通常的 ±1 规范变量的一阶模型已很好地拟合了分子量响应, 得到如下模型:

$$\hat{y} = 2\ 000 + 125x_1 + 40x_2$$

(a) 求最速上升路径.

(b) 希望移动到分子量在 2 500 以上的区域. 利用已从此实验区域内得到的信息, 沿最速上升路径移至所希望的区域大约需要多少步?

11.7 通常在计算最速上升路径时假定模型是真正的一阶模型, 即模型中不含交互作用. 然而, 即使在有交互作用时, 不考虑交互作用的最速上升路径通常仍然可以得到好的结果. 为了说明这个结论, 假设用规范变量 ($-1 \leqslant x_i \leqslant 1$) 得到如下拟合模型:

$$\hat{y} = 20 + 5x_1 - 8x_2 + 3x_1x_2$$

(a) 画出忽略交互作用时的最速上升路径.

(b) 画出模型中有交互作用时的最速上升路径. 将此结果与 (a) 中求得的路径进行比较.

*11.8 下表所示的数据是在优化作为 3 个变量 x_1, x_2, x_3 的函数的晶体生长的实验中所收集得的. 希望 y 的值 (以克为单位的产量) 大一些. 拟合一个二阶模型并分析拟合曲面. 在什么条件下能实现最优的晶体生长?

x_1	x_2	x_3	y	x_1	x_2	x_3	y
−1	−1	−1	66	0	−1.682	0	68
−1	−1	1	70	0	1.682	0	63
−1	1	−1	78	0	0	−1.682	65
−1	1	1	60	0	0	1.682	82
1	−1	−1	80	0	0	0	113
1	−1	1	70	0	0	0	100
1	1	−1	100	0	0	0	118
1	1	1	75	0	0	0	88
−1.682	0	0	100	0	0	0	100
1.682	0	0	80	0	0	0	85

11.9 下列数据是一位化学工程师所收集得的. 响应 y 是渗透时间, x_1 是温度, x_2 是压强. 拟合一个二阶模型.

x_1	x_2	y	x_1	x_2	y
−1	−1	54	0	1.414	51
−1	1	45	0	0	41
1	−1	32	0	0	39
1	1	47	0	0	44
−1.414	0	50	0	0	42
1.414	0	53	0	0	40
0	−1.414	47			

(a) 如果目标是最小化渗透时间, 你会推荐什么运行条件?

(b) 如果目标是使过程在平均渗透时间非常接近 46 处运行, 你会推荐什么运行条件?

11.10 为拟合一个二阶模型, 用正六边形设计进行实验, 收集得下列数据:

x_1	x_2	y	x_1	x_2	y
1	0	68	0	0	58
0.5	$\sqrt{0.75}$	74	0	0	60
−0.5	$\sqrt{0.75}$	65	0	0	57
−1	0	60	0	0	55
−0.5	$-\sqrt{0.75}$	63	0	0	69
0.5	$-\sqrt{0.75}$	70			

(a) 拟合此二阶模型.

(b) 进行正则分析. 曲面属于哪种类型?

(c) 在稳定点处, 运行条件 x_1 和 x_2 取什么值?

(d) 如果目标是使获得的响应尽可能靠近 65, 你会在什么条件下运行此过程?

11.11 实验者已进行了一个 Box-Behnken 设计的实验, 结果如下, 其中响应变量是聚合物的黏度.

水平	温度	搅拌速率	压强	x_1	x_2	x_3
高	200	10.0	25	+1	+1	+1
中	175	7.5	20	0	0	0
低	150	5.0	15	−1	−1	−1

试验	x_1	x_2	x_3	y_1	试验	x_1	x_2	x_3	y_1
1	−1	−1	0	535	9	0	−1	−1	595
2	+1	−1	0	580	10	0	+1	−1	648
3	−1	+1	0	596	11	0	−1	+1	532
4	+1	+1	0	563	12	0	+1	+1	656
5	−1	0	−1	645	13	0	0	0	653
6	+1	0	−1	458	14	0	0	0	599
7	−1	0	+1	350	15	0	0	0	620
8	+1	0	+1	600					

(a) 拟合此二阶模型.

(b) 进行正则分析. 曲面属于哪种类型?

(c) 在稳定点处, 运行条件 x_1, x_2, x_3 取什么值?

(d) 如果获得尽可能靠近 600 的黏度很重要, 你会推荐什么操作条件?

11.12 考虑下表所示的三变量中心复合设计. 假定希望最大化转化率 (y_1), 且使活性 (y_2) 在 55 至 60 之间, 分析数据并得出结论.

试验	时间 (分钟)	温度 (℃)	催化剂 (%)	转化率y_1	活性y_2
1	−1.000	−1.000	−1.000	74.00	53.20
2	1.000	−1.000	−1.000	51.00	62.90
3	−1.000	1.000	−1.000	88.00	53.40
4	1.000	1.000	−1.000	70.00	62.60
5	−1.000	−1.000	1.000	71.00	57.30
6	1.000	−1.000	1.000	90.00	67.90
7	−1.000	1.000	1.000	66.00	59.80
8	1.000	1.000	1.000	97.00	67.80

(续)

试验	时间 (分钟)	温度 (℃)	催化剂 (%)	转化率y_1	活性y_2
9	0.000	0.000	0.000	81.00	59.20
10	0.000	0.000	0.000	75.00	60.40
11	0.000	0.000	0.000	76.00	59.10
12	0.000	0.000	0.000	83.00	60.60
13	−1.682	0.000	0.000	76.00	59.10
14	1.682	0.000	0.000	79.00	65.90
15	0.000	−1.682	0.000	85.00	60.00
16	0.000	1.682	0.000	97.00	60.70
17	0.000	0.000	−1.682	55.00	57.40
18	0.000	0.000	1.682	81.00	63.20
19	0.000	0.000	0.000	80.00	60.80
20	0.000	0.000	0.000	91.00	58.90

11.13 一位切割工具制造者对以小时为单位的刀具寿命 (y_1) 和以美元为单位的刀具成本 (y_2) 研究了两个经验方程. 两个模型都是钢材硬度 (x_1) 和制造时间 (x_2) 的线性函数. 两个方程是

$$\hat{y}_1 = 10 + 5x_1 + 2x_2, \qquad \hat{y}_2 = 23 + 3x_1 + 4x_2$$

两个方程都在范围 $-1.5 \leqslant x_i \leqslant 1.5$ 内有效. 单件刀具成本必须低于 27.5 美元, 寿命必须超过 12 小时, 以使产品具有竞争性. 对此过程存在一组行得通的操作条件吗? 你建议此过程怎样实行?

11.14 进行一项化学汽相淀积过程的中心复合设计的实验, 所得实验数据如下表所示. 对设计的每个试验点, 同时进行 4 个实验单元, 响应是 4 个实验单元中厚度的均值与方差.

x_1	x_2	$\bar{y}$	s^2	x_1	x_2	$\bar{y}$	s^2
−1	−1	360.6	6.689	0	1.414	497.6	7.649
1	−1	445.2	14.230	0	−1.414	397.6	11.740
−1	1	412.1	7.088	0	0	530.6	7.836
1	1	601.7	8.586	0	0	495.4	9.306
1.414	0	518.0	13.130	0	0	510.2	7.956
−1.414	0	411.4	6.644	0	0	487.3	9.127

(a) 对均值响应拟合一个模型, 并进行残差分析.

(b) 对方差响应拟合一个模型, 并进行残差分析.

(c) 对 $\ln(s^2)$ 拟合一个模型, 这个模型比你在 (b) 中得到的模型好吗?

(d) 假设希望平均厚度在区间 450 ± 25 内. 找出可以达到此目标且最小化方差的运行条件.

(e) 讨论在 (d) 中的方差最小化问题. 你已经最小化总的过程方差了吗?

*11.15 证明正交一阶设计也是一阶可旋转设计.

11.16 证明 2^k 设计添加上 n_C 个中心点不影响 β_i ($i = 1, 2, \cdots, k$) 的估计量, 但截距 β_0 的估计量是所有 $2^k + n_C$ 个观测值的平均值.

11.17 **可旋转的中心复合设计**. 如果在 a 或 b (或两者) 是奇数时 $\sum\limits_{u=1}^{n} x_{iu}^{a} x_{ju}^{b}=0$, 且 $\sum\limits_{u=1}^{n} x_{iu}^{4}=3\sum\limits_{u=1}^{n} x_{iu}^{2} x_{ju}^{2}$, 则可证明二阶设计是可旋转的. 证明, 对中心复合设计, 这些条件能推导出 $\alpha = (n_F)^{1/4}$, 其中 n_F 是析因部分的点数.

11.18 证明下面的中心复合设计区组是正交的.

区组 1			区组 2			区组 3		
x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
0	0	0	0	0	0	−1.633	0	0
0	0	0	0	0	0	1.633	0	0
1	1	1	1	1	−1	0	−1.633	0
1	−1	−1	1	−1	1	0	1.633	0
−1	−1	1	−1	1	1	0	0	−1.633
−1	1	−1	−1	−1	−1	0	0	1.633
						0	0	0
						0	0	0

11.19 **中心复合设计的区组化**. 考虑 $k = 4$ 个变量分为两个区组的中心复合设计. 总可以求得正交区组的可旋转的设计吗?

11.20 怎样将正六边形设计分为两个正交区组进行试验?

11.21 一个化工过程的前 4 个循环的产率如下表所示. 变量是: 浓度百分率 (x_1), 其水平为 30、31 和 32; 温度 (x_2), 其水平为 140、142 和 144 °F. 用 EOVP 法分析.

循环	条件				
	(1)	(2)	(3)	(4)	(5)
1	60.7	59.8	60.2	64.2	57.5
2	59.1	62.8	62.5	64.6	58.3
3	56.6	59.1	59.0	62.3	61.1
4	60.5	59.8	64.5	61.0	60.1

11.22 设用阶为 d_1 的模型 (如 $\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$) 来逼近一响应曲面, 真实曲面由阶数为 $d_2 > d_1$ 的模型描述, 也就是, $\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2$.

(a) 证明回归系数是有偏的, 具体说, $\mathrm{E}(\hat{\boldsymbol{\beta}}_1) = \boldsymbol{\beta}_1 + \boldsymbol{A}\boldsymbol{\beta}_2$, 其中 $\boldsymbol{A} = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2$. 通常称 $\boldsymbol{A}$ 为别名矩阵.

(b) 如果 $d_1 = 1$, $d_2 = 2$, 并用完全 2^k 设计来拟合模型, 用 (a) 中的结果来确定别名结构.

(c) 如果 $d_1 = 1$, $d_2 = 2$, $k = 3$, 用 2^{3-1} 设计拟合模型, 求别名结构.

(d) 如果 $d_1 = 1$, $d_2 = 2$, $k = 3$, 并用思考题 11.3 中的单纯形设计拟合模型, 确定别名结构并与 (c) 中的结果进行比较.

*11.23 设需要设计一个实验在指定区域上拟合一个二次模型, 该区域为 $-1 \leqslant x_i \leqslant +1$, $i = 1, 2$, 且满足约束 $x_1 + x_2 \leqslant 1$. 如果破坏约束, 过程将不能正常工作. 可以进行至多不超过 $n = 12$ 次试验. 试建立下列设计:

(a) 中心点在 $x_1 = x_2 = 0$ 的 "内接" CCD.

(b) 中心点在 $x_1 = x_2 = -0.25$ 的 "内接" 3^2 析因设计.

(c) D 最优设计.

(d) 除了重复点都在中心点外, 其余均与 (c) 相同的改进的 D 最优设计.

(e) 计算上述每个设计的 $|(\boldsymbol{X}'\boldsymbol{X})^{-1}|$ 准则.

(f) 计算上述每个设计相对于 (c) 中 D 最优设计的效率.

(g) 你更愿意用哪个设计? 为什么?

*11.24 考虑用 2^3 设计来拟合一个一阶模型.

(a) 计算此设计的 D 准则 $|(\boldsymbol{X}'\boldsymbol{X})^{-1}|$.

(b) 计算此设计的 A 准则 $\mathrm{tr}(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

(c) 求此设计最大的尺度化预测方差. 它是 G 最优的吗?

11.25 对一个含二因子交互作用的一阶模型重做思考题 11.24.

11.26 一化学工程师希望对一新的生产过程拟合一条校准曲线, 用于测量他所在工厂加工的产品中的一种特殊成分的浓度. 可以抽取已知浓度的 12 个样品. 工程师想建立一个用于所测浓度的模型. 他猜想, 对于所测浓度作为已知浓度函数的模型, 线性的校准曲线 (即 $y = \beta_0 + \beta_1 x + \varepsilon$, 其中 x 是实际浓度) 是合适的. 考虑下面 4 个实验设计. 设计 1 由 6 个在已知浓度 1 和 6 个在已知浓度 10 处的试验组成, 设计 2 由在已知浓度 1、5.5 和 10 处各 4 个试验组成, 设计 3 由在已知浓度 1、4、7 和 10 处各 3 个试验组成, 设计 4 由在已知浓度 1 和 10 处各 3 个试验以及 5.5 处 6 个试验组成.

(a) 在同一张图中画出所有 4 个设计浓度在 $1 \leqslant x \leqslant 10$ 范围内的尺度化预测方差. 哪个设计更好?

(b) 对每个设计计算 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 的行列式. 根据 D 准则, 哪个设计更好?

(c) 计算每个设计相对于 (b) 中得到的 "最佳" 设计的 D 效率.

(d) 对于每个设计, 计算在点集 $x = 1, 1.5, 2, 2.5, \cdots, 10$ 上的平均预测方差. 根据 V 准则, 哪个设计更好?

(e) 计算每个设计相对于 (d) 中得到的最佳设计的 V 效率.

(f) 每个设计的 G 效率是什么?

11.27 假定工程师希望拟合的模型是二次模型, 重做思考题 11.26. 显然, 现在只能考虑设计 2、设计 3 和设计 4.

*11.28 一实验者希望进行有 3 个分量的混料实验. 对分量所占比例的约束如下:

$$0.2 \leqslant x_1 \leqslant 0.4, \quad 0.1 \leqslant x_2 \leqslant 0.3, \quad 0.4 \leqslant x_3 \leqslant 0.7$$

(a) 用 D 准则创建一个有 $n = 14$ 个试验、4 次重复的用于拟合二次混料模型的实验.

(b) 画出实验区域.

(c) 用 D 准则创建一个拟合二次混料模型的实验, 它有 $n = 12$ 个试验且其中的 3 个试验是重复试验.

(d) 评述你已得到的两个设计.

*11.29 Myers and Montgomery (2002) 描述了一个含 3 种成分的汽油混合实验. 对混料比例没有限制, 使用有如下 10 个试验点的设计.

设计点	x_1	x_2	x_3	y, 英里/加仑	设计点	x_1	x_2	x_3	y, 英里/加仑
1	1	0	0	24.5, 25.1	6	0	$\frac{1}{2}$	$\frac{1}{2}$	23.5
2	0	1	0	24.8, 23.9	7	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	24.8, 24.1
3	0	0	1	22.7, 23.6	8	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	24.2
4	$\frac{1}{2}$	$\frac{1}{2}$	0	25.1	9	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$	23.9
5	$\frac{1}{2}$	0	$\frac{1}{2}$	24.3	10	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{3}$	23.7

(a) 实验者用的是哪一类设计?

(b) 对上表数据拟合一个二次混料模型. 这个模型合适吗?

(c) 画出响应曲面的等高线图. 为了最大化每加仑英里数, 你推荐用哪种混料配方?

第 12 章　稳健参数设计与过程稳健性研究

本章纲要

12.1　引言
12.2　直积表设计
12.3　直积表设计的分析
12.4　组合表设计及响应模型法
12.5　设计的选择
第 12 章补充材料
S12.1　稳健参数设计的田口方法
S12.2　田口技术方法

12.1　引　　言

稳健参数设计 (Robust Parameter Design, RPD) 是一种产品实现活动的方法, 它在过程或产品中强调选择可控因子 (或参数) 的水平以完成如下两个目标: (1) 保证输出的响应均值是所期望的水平或目标, (2) 保证围绕目标值的变异性尽可能小. 当对一个过程进行 RPD 研究时, 通常称为**过程稳健性研究**. 20 世纪 80 年代由日本工程师田口玄一发展的一般 RPD 问题引入到了美国 (参阅 Taguchi and Wu, 1980, 及 Taguchi, 1987). 田口提出了一种方法, 他用设计过的实验和某些分析实验结果数据的新方法来解决 RPD 问题. 其基本原理和技术方法在工程师和统计学家中引起了广泛关注, 在 20 世纪 80 年代, 他的那套方法被用于许多大企业, 其中包括 AT&T 贝尔实验室、福特汽车公司和施乐. 这些技术在统计和工程领域引发了广泛的争论. 引起争议的并不是基本的 RPD 问题 (虽然它本身其实是极为重要的问题), 而是田口所倡导的实验过程和数据分析方法. 大量的分析表明, 田口的技术方法通常效率很低, 在许多情况下甚至无效. 因此, 随后出现了解决 RPD 问题的新方法的大量研究与发展. 从这些成果中可以看到, 响应曲面方法 (RSM) 显示出它作为解决 RPD 问题的一种方法, 允许我们使用田口的稳健设计的概念, 但 RSM 提供了更好的和更有效的方法进行设计和分析.

本章考虑解决 RPD 问题的 RSM 方法. 有关原来田口的方法的更多信息, 包括指出这个方法的缺陷和低效率性的讨论, 在本章的补充材料中介绍. 其他有用的参考资料有 Box (1988), Box, Bisgaard, and Fung (1988), Hunter (1985,1989), Montgomery (1999), Myers and Montgomery (2002), Pignatiello and Ramberg(1992), 以及 Nair 等人汇编的专题小组的讨论 (1992).

在稳健设计问题中, 重点通常是下面的一个或几个问题.

(1) 设计系统: 一旦系统投入实际使用, 设计系统使得它对影响系统性能的环境因子不敏感. 例如, 开发一种可以在各种气候条件下保持较长寿命的外墙涂料. 因为气候条件并不是完全可以预测的, 当然不是常量, 产品配方设计师希望涂料对影响涂料磨损和光洁度的温度、湿度以及降水量因子在大范围内的变化是稳健的.

(2) 设计产品: 使产品对系统组件所传递的变异性不敏感. 例如, 设计一种电子放大器, 它不受组成系统的晶体管、电阻、电源有关参数变异性的影响, 输出电压尽可能接近所希望的目标值.

(3) 设计过程: 即使有些过程变量 (如温度) 或原材料性能不易被精确控制, 所制造的产品

仍将尽可能接近所希望的目标规格.

(4) 确定过程运行条件：使得关键的过程特征尽可能接近所希望的目标值并在目标值附近的变异最小. 这类问题的实例很多. 例如, 在半导体工厂, 我们希望晶片上氧化物的厚度尽可能接近所希望的目标平均厚度, 且整个晶片上厚度的变化程度 (均匀性的一种度量) 尽可能小.

RPD 问题并不是一个新问题. 数 10 年来, 产品和过程的设计者/开发者一直都在关心稳健性问题, 并且远在田口的成就之前, 就已经在致力于解决这个问题. 用于实现稳健性的典型方法之一是用更坚固的零部件, 或有更小容差的零部件, 或用不同的材料, 重新设计产品. 然而, 这会导致**过度设计**(overdesign) 问题, 从而使得产品价格更高, 更难制造, 或重量难以承受. 有时, 可以利用不同的设计方法, 或采用新技术. 例如, 过去许多年里, 汽车速度计由金属缆线传导, 缆线中的润滑剂性能随着时间逐渐退化, 这可能导致在寒冷气候条件下运行有噪声或车速测量不稳定. 有时缆线会破裂, 产生昂贵的修理费用. 这是一个由产品老化引起的稳健性问题的例子. 而现代的汽车用电子速度计, 不会出现上述问题. 在一个过程环境中, 老的设备可以用能改进过程稳健性的新工具取代, 但通常成本很大. 另一个可能性是对影响稳健性的变量施行更严格的控制. 例如, 如果环境条件的波动引发稳健性问题, 那么这些条件或许必须被更严格地控制. 半导体厂中利用除尘室就是为了控制环境条件. 有些情况下, 会对影响稳健性的原材料的性能或过程变量进行更严格的控制. 尽管这些经典的方法仍然有用, 但田口的主要贡献是认识到实验的设计和其他统计工具在许多情况下可以用于解决这个问题.

田口方法的一个重要观点是, 某类变量能影响系统重要响应变量的变异. 我们称这类变量是**噪声变量**或**不可控变量**. 前面已讨论过这个概念 (比如, 见图 1.1). 这些噪声因子通常是诸如温度或相对湿度之类环境条件的函数, 也可能是原材料各批次间或持续的过程中性能的变化, 还可能是难以控制或保持在指定目标值的过程变量. 有时, 它们可能涉及用户操作或使用产品的方式. 噪声变量通常在研究或开发阶段是可控的, 但在生产或使用阶段是不可控的. RPD 问题的主要工作就是识别出影响过程或产品性能的可控变量和噪声变量, 并找到可控变量的设置, 来最小化由噪声变量引起的变异.

为了说明可控变量和噪声变量, 考虑一名正在设计蛋糕粉配方的产品开发人员. 开发人员必须详细指定蛋糕粉的成分及混合物, 包括面粉、糖、奶粉、氢化油、玉米淀粉以及香料的用量. 在生产蛋糕粉时, 容易控制这些变量. 当用户烘焙蛋糕时, 需要加水, 并将湿的和干的成分配料混合成蛋糕面糊, 在烤箱中以指定的温度和时间烘焙. 但产品的配方师并不能精确控制用户究竟向干蛋糕粉中加多少水, 湿的和干的成分混合得有多好, 精确的烘焙时间与精确的烤箱温度. 一般可以指定这些变量的目标值, 但它们的确是噪声变量, 因为不同的用户使用这些因子的水平存在差异 (或许是很大的差异). 因此, 产品配方师遇到了一个稳健设计问题. 目标是设计出好的蛋糕粉配方, 不管最后做蛋糕时噪声变量如何变化, 做出的蛋糕都很好, 可以满足或超过用户的期望.

12.2 直积表设计

对于 RPD 问题, 原来的田口方法以可控变量的统计设计和噪声变量的统计设计为中心进行. 这两个设计被 "直积", 即, 将可控变量设计中的每一个处理组合与噪声变量设计的每个处理组合结合在一起进行试验. 这类实验设计就称为**直积表设计**(crossed array design).

我们用思考题 8.7 中的簧片实验来说明直积表设计方法. 在这个实验中, 研究 5 个因子对汽车上使用的簧片自由高度的效应. 实验中的 5 个因子是: A = 炉温, B = 加热时间, C = 传递时间, D = 保持时间, E = 淬火油温. 这原本就是一个 RPD 问题, 淬火油温是噪声变量. 此实验的数据见表 12.1. 可控因子的设计是生成元为 $D = ABC$ 的 2^{4-1} 分式析因设计. 称此设计为**内表**(inner array) 设计. 单个噪声因子的设计用 2^1 设计, 称为**外表**(outer array) 设计. 注意外表的每个试验是如何完成内表中所有的 8 个处理组合的试验, 产生直积表结构的. 簧片实验中, 在 16 个不同的设计点的每个点上进行 3 次重复试验, 实验结果是自由高度的 48 个观测值.

表 12.1 簧片实验

A	B	C	D	$E = -$	$E = +$	$\bar{y}$	s^2
−	−	−	−	7.78, 7.78, 7.81	7.50, 7.25, 7.12	7.54	0.090
+	−	−	+	8.15, 8.18, 7.88	7.88, 7.88, 7.44	7.90	0.071
−	+	−	+	7.50, 7.56, 7.50	7.50, 7.56, 7.50	7.52	0.001
+	+	−	−	7.59, 7.56, 7.75	7.63, 7.75, 7.56	7.64	0.008
−	−	+	+	7.54, 8.00, 7.88	7.32, 7.44, 7.44	7.60	0.074
+	−	+	−	7.69, 8.09, 8.06	7.56, 7.69, 7.62	7.79	0.053
−	+	+	−	7.56, 7.52, 7.44	7.18, 7.18, 7.25	7.36	0.030
+	+	+	+	7.56, 7.81, 7.69	7.81, 7.50, 7.59	7.66	0.017

直积表设计的一个重点是, 它提供了可控因子与噪声因子之间交互作用的信息. 这些交互作用对解决 RPD 问题是极为重要的. 例如, 考虑图 12.1 中二因子交互作用的图形, 其中 x 是可控因子, z 是噪声因子. 图 12.1a 中, x 与 z 之间没有交互作用, 因此, 不存在影响噪声因子 z 的变异性传递给响应变异性的可控变量 x 的设置. 然而, 在图 12.1b 中, x 与 z 之间有强烈的交互作用. 当 x 处于低水平时, 响应变量的变异性比因 x 处于高水平时低得多. 于是, 除非有至少一个可控因子与噪声因子有交互作用, 否则不存在稳健设计问题. 随后我们会看到, 将重点放在识别这些交互作用并建立相应模型是解决 RPD 问题高效实用方法的关键之一.

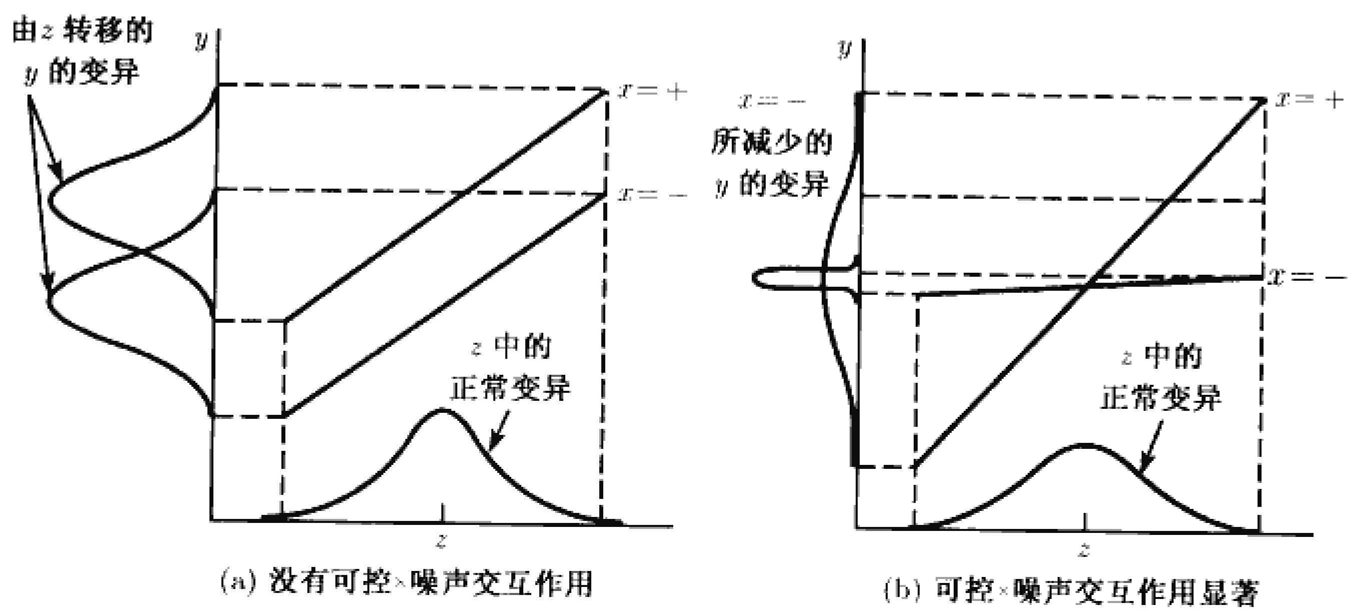


图 12.1 在稳健设计中的可控 × 噪声交互作用

表 12.2 给出了 RPD 问题的又一个例子, 这个例子来自 Byrne and Taguchi (1987). 该问题涉及一种弹性塑料连接器的开发, 该连接器与尼龙管装配在一起, 要求它能提供所需的拉开

力. 有 4 个三水平的可控因子 (A = 障碍物, B = 连接器壁厚, C = 嵌入深度, D = 粘合剂量), 3 个二水平的噪声或不可控因子 (E = 规定时间, F = 规定温度, G = 规定相对湿度). 表 12.2 的面板 (a) 中包含了可控因子的内表设计. 这个设计是三水平分式析因设计, 3^{4-2} 设计. 表 12.2 的面板 (b) 中包含了噪声因子的 2^3 外表设计. 像前面一样, 现在, 内表中每一个试验点要进行外表的所有处理组合的试验, 产生表中所示的、有 72 个拉开力观测的直积表设计.

考察表 12.2 中的直积表设计, 可以看到田口设计策略中的主要问题, 即, 直积表方法会导致非常大的实验. 在我们的例子中, 只有 7 个因子, 而这个设计有 72 个试验点. 此外, 内表设计是分辨度为 III 的 3^{4-2} 设计 (见第 9 章对此设计的讨论), 所以, 尽管已做了那么多个试验, 仍然得不到可控因子之间的交互作用的任何信息. 实际上, 即使是有关主效应的信息, 也可能受到破坏, 因为主效应与二因子交互作用严重混杂. 12.4 节将介绍**组合表设计**, 一般而言, 它比直积表设计有效得多.

表 12.2 连接器拉开力实验的设计

					(b) 外表							
				E	−	−	−	−	+	+	+	+
				F	−	−	+	+	−	−	+	+
				G	−	+	−	+	−	−	−	+
		(a) 内表										
试验	A	B	C	D								
1	−1	−1	−1	−1	15.6	9.5	16.9	19.9	19.6	19.6	20.0	19.1
2	−1	0	0	0	15.0	16.2	19.4	19.2	19.7	19.8	24.2	21.9
3	−1	+1	+1	+1	16.3	16.7	19.1	15.6	22.6	18.2	23.3	20.4
4	0	−1	0	+1	18.3	17.4	18.9	18.6	21.0	18.9	23.2	24.7
5	0	0	+1	−1	19.7	18.6	19.4	25.1	25.6	21.4	27.5	25.3
6	0	+1	−1	0	16.2	16.3	20.0	19.8	14.7	19.6	22.5	24.7
7	+1	−1	+1	0	16.4	19.1	18.4	23.6	16.8	18.6	24.3	21.6
8	+1	0	−1	+1	14.2	15.6	15.1	16.8	17.8	19.6	23.2	24.2
9	+1	+1	0	−1	16.1	19.9	19.3	17.3	23.1	22.7	22.6	28.6

12.3 直积表设计的分析

田口提出了直积表实验中汇总数据的两个统计量: 内表每试验点所对应的外表所有试验数据的平均值, 以及称之为**信噪比**(signal-to-noise ratio) 的试图组合均值与方差信息的一个综合性统计量. 定义这些信噪比据称是为了使信噪比的最大值最小化由噪声变量所传递的变异. 然后进行分析以确定可控因子的哪些设置可以使得: (1) 均值尽可能接近所希望的目标值, (2) 信噪比取得最大值. 信噪比是一个有问题的统计量: 它们可能导致位置效应和散度效应的混淆, 通常也不能由此找到解决最小化传递变异的 RPD 问题的期望方案. 这个问题将在本章的补充材料中详细讨论.

对直积表设计更合适的分析是直接对响应的均值与方差建立模型, 其中, 内表上每一观测的样本均值和样本方差由外表中所有试验的结果计算得到. 根据直积表的结构, 样本均值 $\bar{y}_i$ 与样本方差 s_i^2 是在噪声变量取相同水平时计算得到的, 所以这些量之间的任何差异都是由可控变量

水平之间的差异造成的. 因此, 选择优化均值同时最小化方差的可控变量的水平是一种有用的方法.

为了说明这种方法, 考虑表 12.1 中的簧片实验. 表中最后两列对内表的每一试验列出了样本均值 $\bar{y}_i$ 与样本方差 s_i^2. 图 12.2 是平均自由高度响应效应的半正态概率图. 显然, 因子 A, B, D 是重要因子. 由于这些因子与三因子交互作用混杂, 有理由得到这些效应是确实存在的结论. 平均自由高度响应的模型是

$$\hat{y}_i = 7.63 + 0.12x_1 - 0.081x_2 + 0.044x_4$$

其中, 各 x 表示原设计因子 A, B, D. 因为样本方差不服从正态分布 (服从尺度化卡方分布), 所以通常最好是分析方差的自然对数. 图 12.3 是 $\ln(s_i^2)$ 响应的效应的半正态概率图. 只有因子 B 是显著的. $\ln(s_i^2)$ 响应的模型是

$$\widehat{\ln(s_i^2)} = -3.74 - 1.09x_2$$

图 12.4 是因子 A 与因子 B 在因子 $D = 0$ 时平均自由高度的等高线图, 图 12.5 是方差响应在原始尺度下的图. 显然, 自由高度的方差随着加热时间 (因子 B) 的增加而减少.

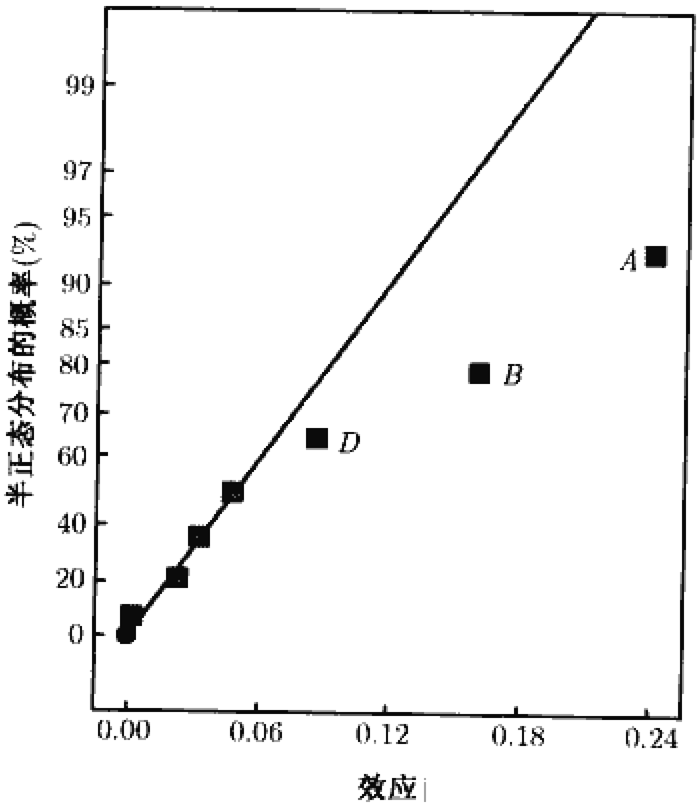


图 12.2　平均自由高度响应的效应的半正态图

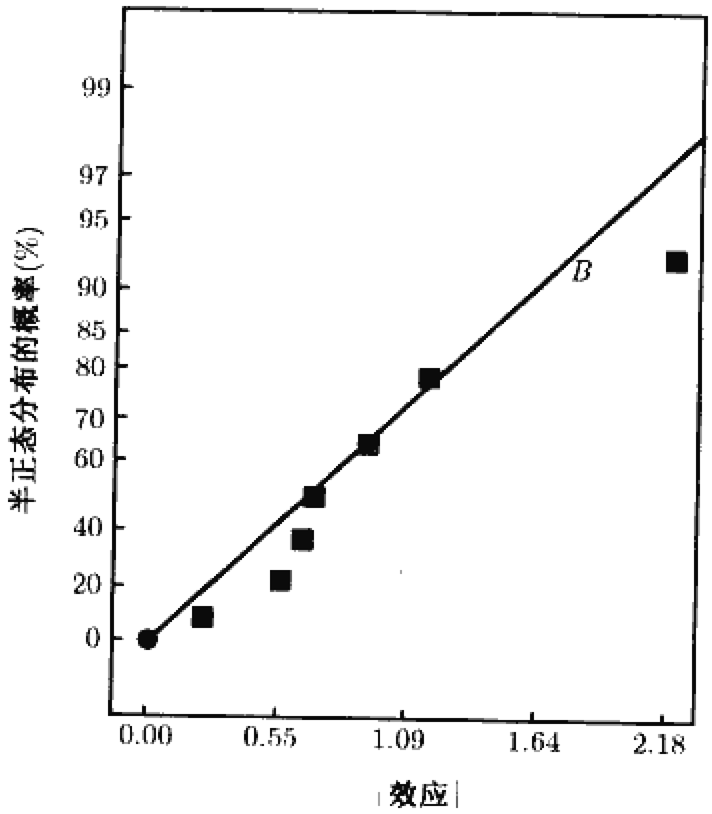


图 12.3　$\ln(s_i^2)$ 响应效应的半正态图

假定实验的目的是找出这样一组运行条件: 它们导致平均自由高度在 7.74 英寸与 7.76 英寸之间, 且可以最小化变异性. 这个标准的多重响应优化问题可以用第 11 章所描述的解决这类问题的任一方法来求解. 图 12.6 是两个响应的重叠等高线图, 其中因子 D = 保持时间在高水平上保持不变. 还可以选择因子 A = 温度在高水平, 因子 B = 加热时间在 0.5 (规范单位), 从而把平均自由高度控制在所希望的界限内, 而方差近似为 0.013 8.

使用直积表设计对均值和方差进行建模的一个缺点是, 它不能直接利用可控变量与噪声变量之间的交互作用. 在某些情况下, 它甚至可能掩饰这些关系. 此外, 方差响应很可能与可控变量有非线性关系 (例如, 参见图 12.5), 这可能使建模过程变得复杂. 12.4 节将介绍克服这些问题的另一种设计策略和建模方法.

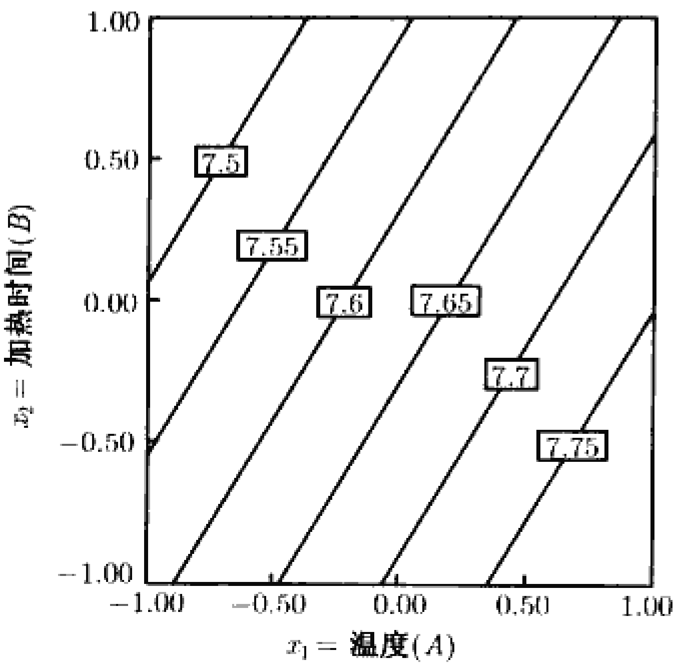


图 12.4 D = 保持时间 = 0 时平均自由高度响应的等高线图

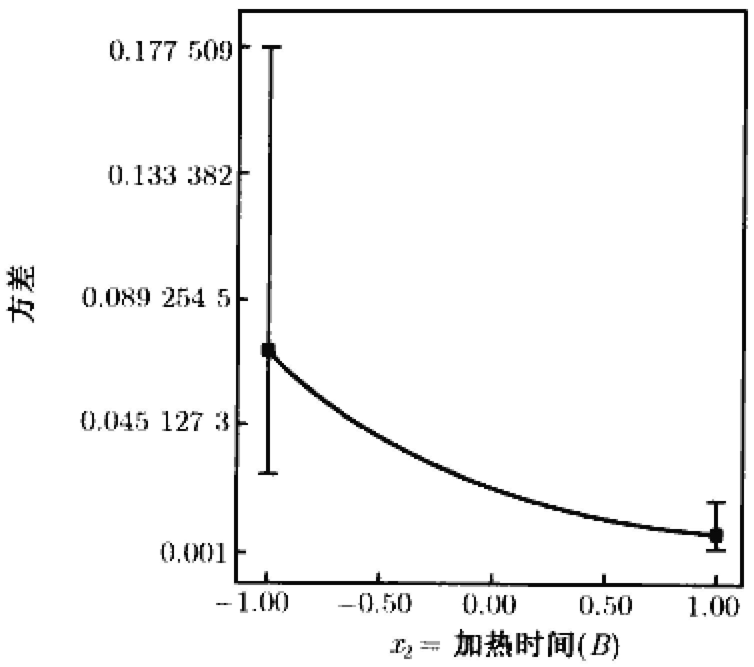


图 12.5 自由高度的方差与 x = 加热时间 (B) 的关系图

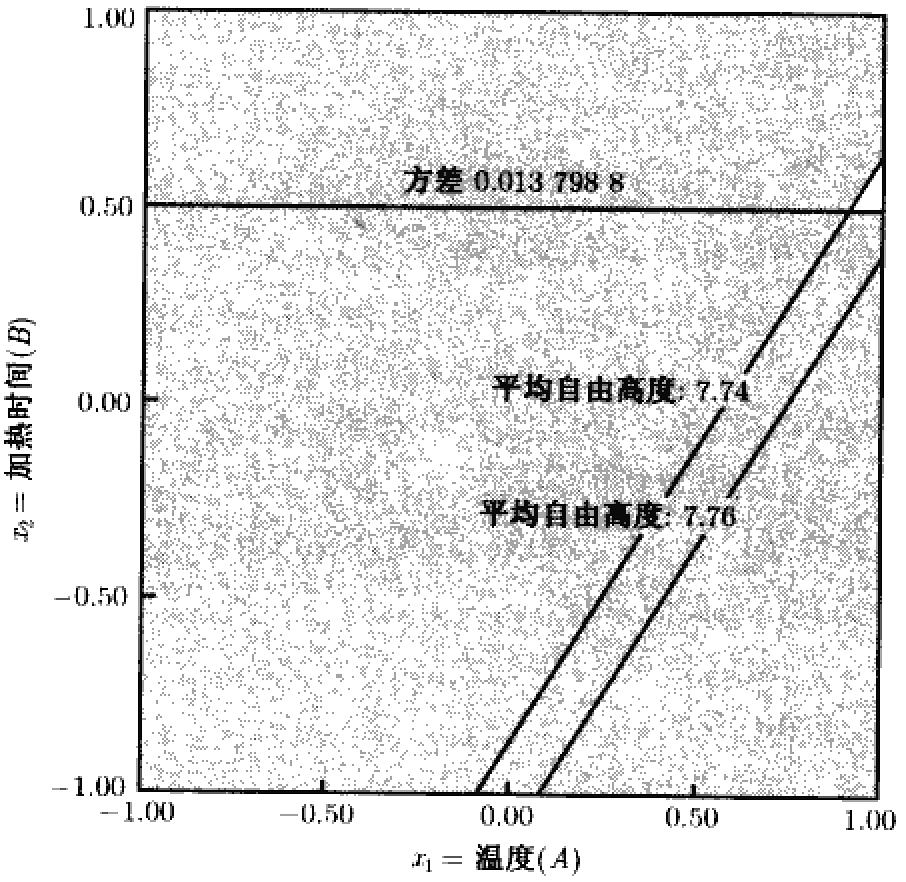


图 12.6 x_4 = 保持时间 (D) 在高水平时平均自由高度和自由高度的方差的重叠等高线图

12.4 组合表设计及响应模型法

如 12.3 节所指出的, 可控因子与噪声因子之间的交互作用是稳健设计问题的关键. 因此, 对于响应使用包括可控因子和噪声因子及其交互作用的**模型**是合理的. 为了说明这一点, 设有两个可控因子 x_1 和 x_2 与一个噪声因子 z_1. 假定可控因子和噪声因子都按通常的规范变量 (即中心在零, 上下限是 $\pm a$) 的形式表达. 若希望使用包含可控变量的一阶模型, 则合理的模型是

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \gamma_1 z_1 + \delta_{11} x_1 z_1 + \delta_{21} x_2 z_1 + \varepsilon \tag{12.1}$$

这个模型包括了可控因子的主效应及其交互作用、噪声变量的主效应、可控变量与噪声变量的

两个交互作用. 这类同时包含可控变量和噪声变量的模型通常称为**响应模型**. 除非回归系数 δ_{11} 和 δ_{21} 中至少有一个不为零, 否则它就不是稳健设计问题.

响应模型方法的重要优势是, 可控因子和噪声因子可以同时放在一个实验设计中, 也就是说, 可以避免使用田口方法的内表和外表结构. 通常称同时包含可控因子和噪声因子的设计为**组合表设计**(combined array design).

与前面提到的相同, 尽管噪声变量对实验目的是可控的, 但我们仍假定它们是随机变量. 我们特地假定噪声变量以规范变量形式表示, 期望为零, 方差为 σ_z^2; 如果噪声变量不止一个, 则假定它们之间的协方差为零. 在这些假定下, 对 (12.1) 式求 y 的期望就易得到响应均值模型. 于是有

$$\mathrm{E}_z(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 \tag{12.2}$$

其中的期望算子的下标 z 表示是对 (12.1) 式中的两个随机变量 z_1 和 ε 两者求期望. 为了求模型中响应 y 的方差, 我们用**误差传递法**. 首先, 将 12.1 式中的响应模型在 $z_1=0$ 处以一阶泰勒级数方法展开. 得

$$\begin{aligned} y &\cong y_{z=0} + \frac{\mathrm{d}y}{\mathrm{d}z_1}(z_1 - 0) + R + \varepsilon \\ &\cong \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + (\gamma_1 + \delta_{11} x_1 + \delta_{21} x_2) z_1 + R + \varepsilon \end{aligned}$$

其中, R 是泰勒级数的余项. 通常情况下将忽略此余项. 现在, y 的方差可以通过对最后一个表达式 (不含 R) 取方差获得. 所得的方差模型是

$$\mathrm{V}_z(y) = \sigma_z^2(\gamma_1 + \delta_{11} x_1 + \delta_{21} x_2)^2 + \sigma^2 \tag{12.3}$$

在方差算子上又用了下标 z, 表示 z_1 和 ε 二者都是随机变量.

(12.2) 式和 (12.3) 式是响应变量的均值和方差的简单模型. 注意以下几点.

(1) 均值模型和方差模型**只包含可控变量**. 这意味着可以通过设置可控变量以使均值达到目标值, 并最小化由噪声变量传递的变异性.

(2) 尽管方差模型只包含可控变量, 但它也包含可控变量与噪声变量*交互作用的回归系数*. 它表明噪声变量是如何影响响应的.

(3) 方差模型是可控变量的**二次函数**.

(4) 方差模型 (除 σ^2 外) 恰好是所拟合响应模型在噪声变量方向上斜率的平方.

为了使用这些模型, 我们要

(1) 进行一项实验, 并拟合一个适当的响应模型, 如 (12.1) 式.

(2) 用响应模型中系数的最小二乘估计取代均值模型和方差模型中的未知回归系数, 用拟合响应模型中的残差均方取代方差模型中的 σ^2.

(3) 用 11.3.4 节中讨论的标准多重响应优化方法优化均值模型和方差模型.

可以很容易地把这些结果一般化. 假设有 k 个可控变量, r 个噪声变量. 包含这些变量的一般模型可以写成

$$y(\boldsymbol{x}, \boldsymbol{z}) = f(\boldsymbol{x}) + h(\boldsymbol{x}, \boldsymbol{z}) + \varepsilon \tag{12.4}$$

其中 $f(\boldsymbol{x})$ 是模型仅包含可控变量的部分, $h(\boldsymbol{x}, \boldsymbol{z})$ 包括了噪声因子的主效应和可控因子与噪声因子的交互作用的项. $h(\boldsymbol{x}, \boldsymbol{z})$ 的典型结构是

$$h(\boldsymbol{x},\boldsymbol{z})=\sum_{i=1}^{r}\gamma_i z_i+\sum_{i=1}^{k}\sum_{j=1}^{r}\delta_{ij}x_i z_j$$

$f(\boldsymbol{x})$ 的结构取决于实验者认为适当的可控变量的模型类型. 合理的选择是带有交互作用的一阶模型和二阶模型. 如果假定噪声变量的均值为零, 方差为 $\sigma_{z_i}^2$, 协方差为零, 且噪声变量与随机误差 ε 的协方差也是零, 则响应的均值模型是

$$\mathrm{E}_z[y(\boldsymbol{x},\boldsymbol{z})]=f(\boldsymbol{x}) \tag{12.5}$$

响应的方差模型是

$$\mathrm{V}_z[y(\boldsymbol{x},\boldsymbol{z})]=\sum_{i=1}^{r}\left[\frac{\partial y(\boldsymbol{x},\boldsymbol{z})}{\partial z_i}\right]^2\sigma_{z_i}^2+\sigma^2 \tag{12.6}$$

Myers and Montgomery (2002) 直接对响应模型应用条件方差算子, 给出了比 (12.6) 式更一般化的形式.

例 12.1 为了说明前面介绍的方法, 重新考虑例 6.2, 它在 2^4 析因设计中研究 4 个因子对化学产品渗透率的效应. 假定因子 A(温度), 在大规模生产时可能是难以控制的, 但在实验期间 (在小试验工厂进行) 是可控的. 其他 3 个因子 [压强 (B)、浓度 (C) 和搅拌速度 (D)], 是容易控制的. 于是, 噪声因子 z_1 是温度, 可控变量 x_1, x_2, x_3 分别是压强、浓度和搅拌速度. 因为可控因子和噪声因子都在同一个设计中, 用于这个实验的 2^4 析因设计是**组合表设计**的一个例子.

由例 6.2 中的结果知, 响应模型是

$$\begin{aligned}\hat{y}(\boldsymbol{x},z_1)&=70.06+\left(\frac{21.625}{2}\right)z_1+\left(\frac{9.875}{2}\right)x_2+\left(\frac{14.625}{2}\right)x_3\\&\quad-\left(\frac{18.125}{2}\right)x_2z_1+\left(\frac{16.625}{2}\right)x_3z_1\\&=70.06+10.81z_1+4.94x_2+7.31x_3-9.06x_2z_1+8.31x_3z_1\end{aligned}$$

由 (12.5) 式和 (12.6) 式, 可求得均值模型和方差模型分别为

$$\mathrm{E}_z[y(\boldsymbol{x},z_1)]=70.06+4.94x_2+7.31x_3$$

和

$$\begin{aligned}\mathrm{V}_z[y(\boldsymbol{x},z_1)]&=\sigma_z^2(10.81-9.06x_2+8.31x_3)^2+\sigma^2\\&=\sigma_z^2(116.91+82.08x_2^2+69.06x_3^2-195.88x_2+179.66x_3-150.58x_2x_3)+\sigma^2\end{aligned}$$

现在假定噪声变量温度的低水平和高水平运行在它的典型值或平均值两边一个标准差, 则 $\sigma_z^2=1$, $\hat{\sigma}^2=19.51$ (这是拟合响应模型所得的残差均方和). 因此, 方差模型变成

$$\mathrm{V}_z[y(\boldsymbol{x},z_1)]=136.42-195.88x_2+179.66x_3-150.58x_2x_3+82.08x_2^2+69.06x_3^2$$

图 12.7 是由 Design-Expert 软件包给出的均值模型的响应等高线图. 为了构造这张图, 我们将噪声因子 (温度) 保持在零, 不显著的可控因子 (压强) 也保持在零. 注意随着浓度和搅拌速度的增加, 平均渗透率是增加的. Design-Expert 也将自动构造方差**平方根**的等高线图, 记为**误差传播**(propagation of error, POE). 显然, 作为可控变量函数的 POE 恰好是响应中传递变异性的标准差. 图 12.8 显示了由 Design-Expert 得到的 POE 的等高线图和三维响应曲面图. (在这张图中, 像前面一样, 噪声变量保持为常数零.)

假设实验者想将平均渗透率保持在 75 左右, 并最小化围绕该值的变异性. 图 12.9 显示了作为显著的可控变量 (浓度和搅拌速度) 函数的平均渗透率与 POE 的重叠等高线图. 为了达到理想的目标, 必须将浓度保持在高水平, 搅拌速度保持在中等水平.

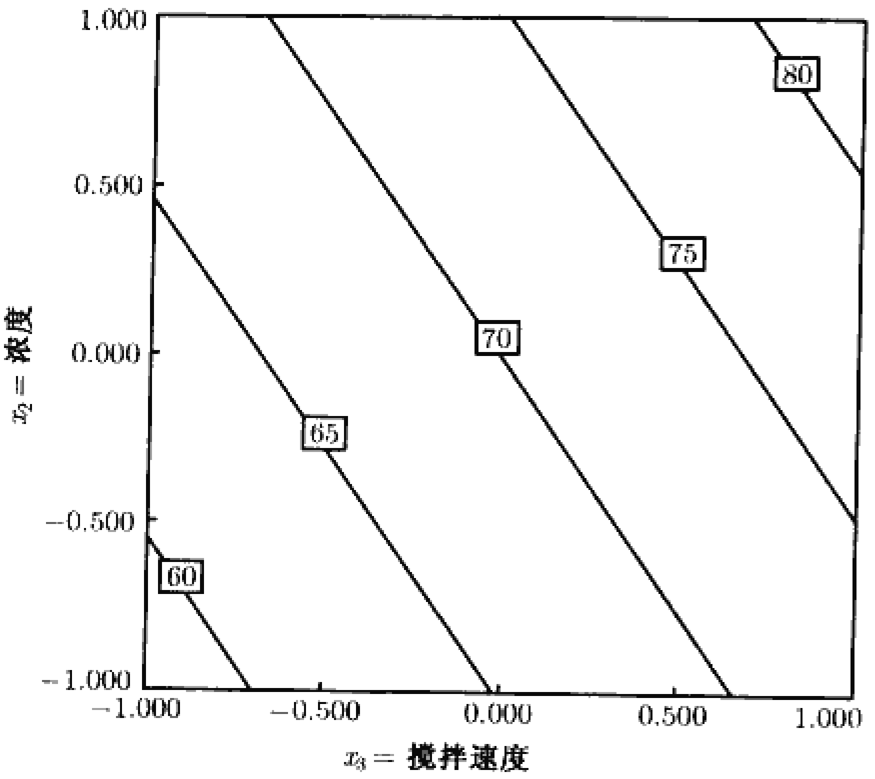
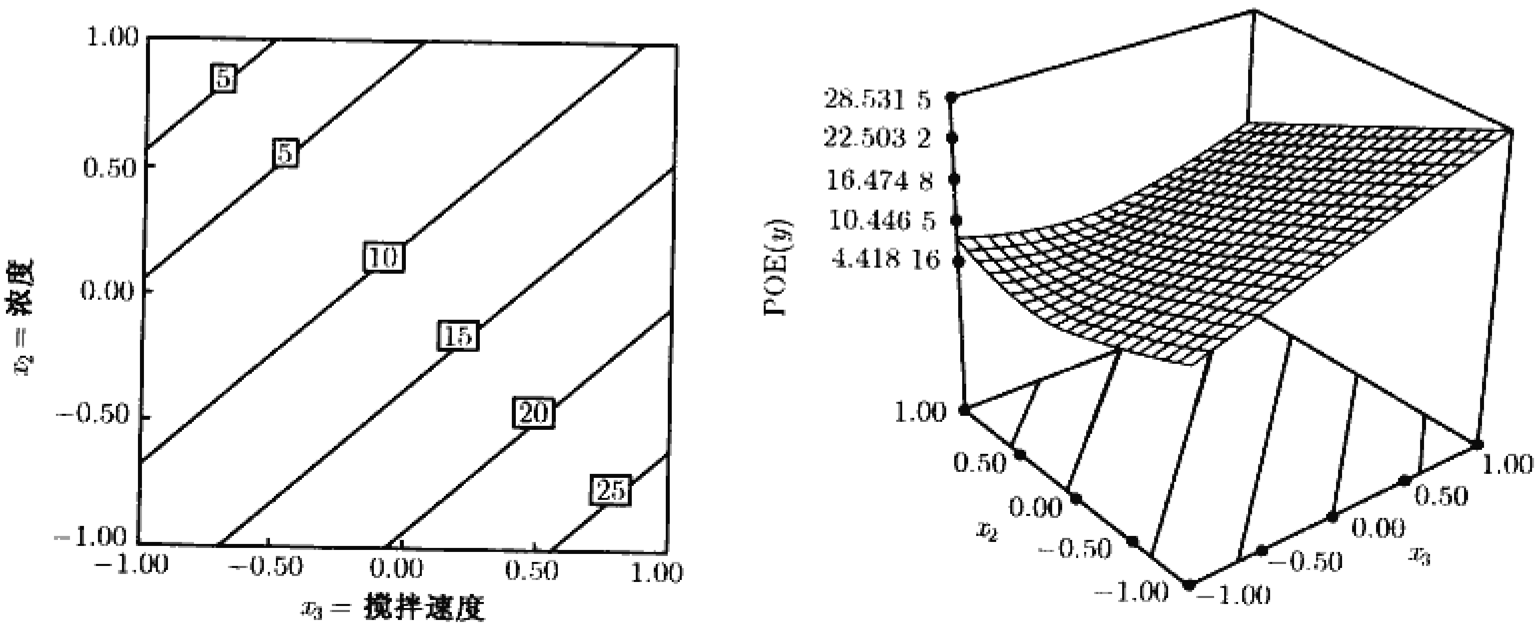


图 12.7 例 12.1 中 z_1 = 温度 = 0 时平均渗透率的等高线图



(a) 等高线图 (b) 响应曲面图

图 12.8 例 12.1 中 z_1 = 温度 = 0 时误差传播的等高线图和响应曲面图

从例 12.1 中可以看到, 渗透率响应的标准差仍然很大. 这说明有时候过程的稳健性研究并不能产生一个完全令人满意的结果. 有必要用其他方法来实现更好的过程性能, 如可以在大规模生产中更精确地控制温度.

例 12.1 给出了用含交互作用的一阶模型 $f(\boldsymbol{x})$ 作为可控因子的模型. 下面给出一个含有二阶模型的例子, 它来自 Montgomery (1999).

例 12.2 半导体工厂进行一项包括 2 个可控变量和 3 个噪声变量的实验. 实验者使用的组合表设计如表 12.3 所示. 这个设计有 23 次试验, 是以 5 个因子的标准 CCD 设计 (立方部分是 2^{5-1} 分式设计) 为基础, 删去了与 3 个噪声因子有关的轴向试验点的中心复合设计的变形. 这个设计支持的响应模型中包括了可控变量的二阶模型的项、3 个噪声变量的主效应以及可控因子和噪声因子之间的交互作用.

拟合所得的响应模型是

$$
\begin{aligned}
\hat{y}(\boldsymbol{x}, \boldsymbol{z}) = &\,30.37 - 2.92x_1 - 4.13x_2 + 2.60x_1^2 + 2.18x_2^2 + 2.87x_1x_2 + 2.73z_1 - 2.33z_2 \\
&+2.33z_3 - 0.27x_1z_1 + 0.89x_1z_2 + 2.58x_1z_3 + 2.01x_2z_1 - 1.43x_2z_2 + 1.56x_2z_3
\end{aligned}
$$

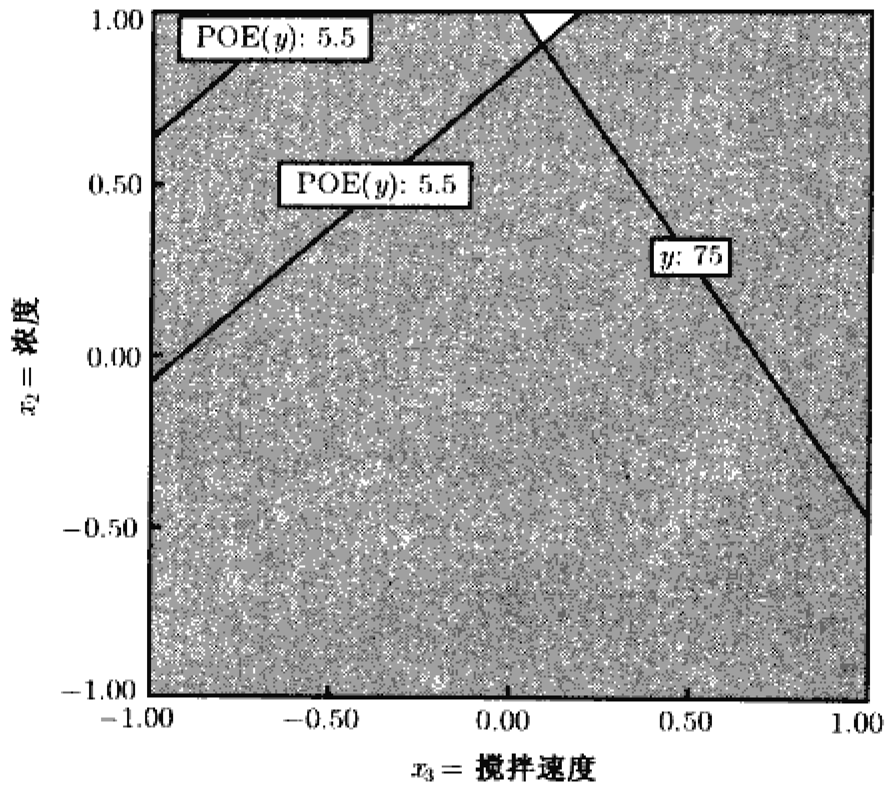


图 12.9　例 12.1 中 z_1= 温度 =0 时渗透率的均值与 POE 的重叠等高线图

表 12.3　例 12.2 中含 2 个可控变量和 3 个噪声变量的组合表实验

试验号	x_1	x_2	z_1	z_2	z_3	y
1	−1.00	−1.00	−1.00	−1.00	1.00	44.2
2	1.00	−1.00	−1.00	−1.00	−1.00	30.0
3	−1.00	1.00	−1.00	−1.00	−1.00	30.0
4	1.00	1.00	−1.00	−1.00	1.00	35.4
5	−1.00	−1.00	1.00	−1.00	−1.00	49.8
6	1.00	−1.00	1.00	−1.00	1.00	36.3
7	−1.00	1.00	1.00	−1.00	1.00	41.3
8	1.00	1.00	1.00	−1.00	−1.00	31.4
9	−1.00	−1.00	−1.00	1.00	−1.00	43.5
10	1.00	−1.00	−1.00	1.00	1.00	36.1
11	−1.00	1.00	−1.00	1.00	1.00	22.7
12	1.00	1.00	−1.00	1.00	−1.00	16.0
13	−1.00	−1.00	1.00	1.00	1.00	43.2
14	1.00	−1.00	1.00	1.00	−1.00	30.3
15	−1.00	1.00	1.00	1.00	−1.00	30.1
16	1.00	1.00	1.00	1.00	1.00	39.2
17	−2.00	0.00	0.00	0.00	0.00	46.1
18	2.00	0.00	0.00	0.00	0.00	36.1
19	0.00	−2.00	0.00	0.00	0.00	47.4
20	0.00	2.00	0.00	0.00	0.00	31.5
21	0.00	0.00	0.00	0.00	0.00	30.8
22	0.00	0.00	0.00	0.00	0.00	30.7
23	0.00	0.00	0.00	0.00	0.00	31.0

均值模型和方差模型分别为

$$\mathrm{E}_{\boldsymbol{z}}[y(\boldsymbol{x},\boldsymbol{z})] = 30.37 - 2.92x_1 - 4.13x_2 + 2.60x_1^2 + 2.18x_2^2 + 2.87x_1x_2$$

和

$$\mathrm{V}_{\boldsymbol{z}}[y(\boldsymbol{x},\boldsymbol{z})] = 19.26 + 6.40x_1 + 24.91x_2 + 7.52x_1^2 + 8.52x_2^2 + 4.42x_1x_2$$

在此, 我们将拟合到的响应模型的参数估计代入均值模型与方差模型, 并且与前面的例子相同, 假定 $\sigma_z^2 = 1$. (由 Design-Expert 得到的) 图 12.10 与图 12.11 分别给出了这些模型的过程均值与 POE (POE 是方差响应曲面的平方根) 的等高线图.

在这个问题中, 较为理想的情况是将过程的均值保持在 30 以下. 由图 12.10 和图 12.11 可以看到, 如果我们希望过程方差较小, 显然必须要有某种折衷. 因为只有两个可控变量, 完成这个折衷的合理方法应该是如图 12.12 所示, 重叠均值响应和方差的等高线. 这张图显示了过程均值小于等于 30, 过程标准差小于等于 5 的等高线. 这两条等高线所围区域是低平均值响应和低过程方差的典型的运行区域.

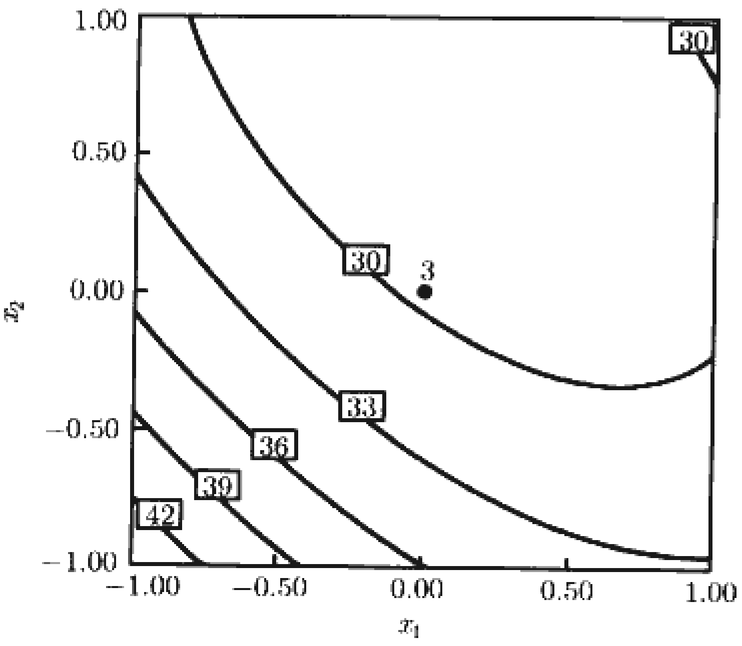


图 12.10 例 12.2 中均值模型的等高线图

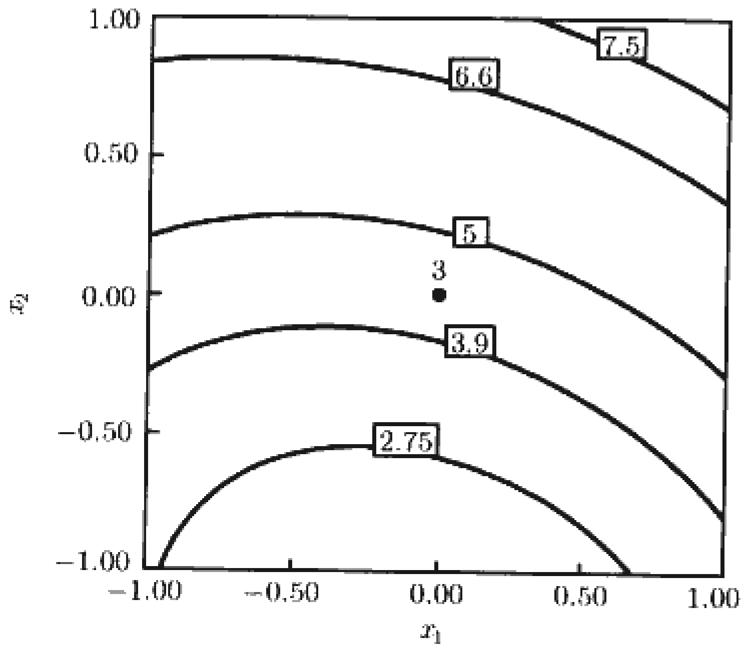


图 12.11 例 12.2 中 POE 的等高线图

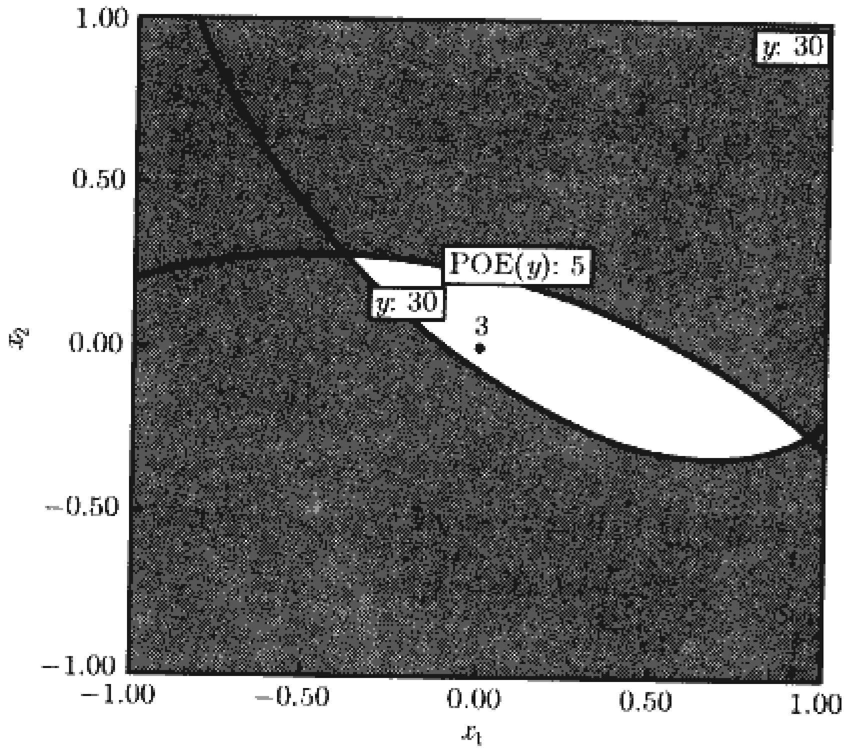


图 12.12 例 12.2 中均值与 POE 的重叠等高线图, 未遮盖的区域表示均值与方差都令人满意时的过程操作条件

12.5 设计的选择

实验设计的选择是 RPD 问题中一个非常重要的方面. 一般而言, 组合表方法得到的设计比直积表得到的设计小. 而且, 响应建模方法允许直接加入可控因子与噪声因子的交互作用, 这通

常优于直接的均值与方差建模. 因此, 本节仅评述组合表设计.

如果所有设计因子都是二水平的, 则对 RPD 研究来说, 分辨度为 V 的设计是一个很好的选择, 因为在假定所有三因子交互效应和更高阶交互效应可以忽略时, 它可以估计所有主效应和二因子交互效应. 在有些情况下, 标准的 2_{V}^{k-p} 分式析因设计是好的选择. 例如, 有 5 个因子, 设计需要 16 次试验. 然而, 有 6 个或更多的因子时, 标准的 2_{V}^{k-p} 分式析因设计相当大. 如第 8 章所述, Design-Expert 软件包包含了更小的分辨度为 V 的二水平设计. 表 12.4 所示的是来自这

表 12.4　有 7 个因子、30 次试验的分辨度为 V 的设计

A	B	C	D	E	F	G
−	+	−	−	−	+	−
+	−	−	−	−	−	−
−	+	+	+	−	+	−
+	−	+	−	+	−	−
−	+	−	+	−	−	+
+	−	+	+	−	−	−
+	+	+	+		+	+
−	+	−	+	+	−	−
+	−	+	−	−	−	+
−	−	−	+	+	−	+
+	−	+	−	−	+	−
−	+	+	+	+	−	+
−	−	+	−	−	−	−
−	−	+	+	−	+	+
+	+	−	+	−	−	−
+	+	−	+	+	−	+
+	+	−	−	+	+	−
−	−	+	+	+	−	−
+	+	+	−	−	−	−
−	+	+	−	−	+	+
−	−	−	−	−	−	+
−	+	+	−	+	+	−
−	+	−	−	+	−	+
−	+	−	+	+	+	+
−	−	−	−	+	+	−
+	−	−	+	−	+	+
+	−	−	−	+	+	+
+	+	+	−	+	+	+
+	−	−	+	+	+	−
+	−	+	+	+	+	+

个软件包的有 7 个因子的设计, 需要 30 次试验. 这个设计可容纳 7 个因子, 它们可以是可控变量与噪声变量的任意组合, 该设计可以估计所有主效应和二因子交互效应.

有时可以用只有很少试验点的设计. 例如, 假设有 3 个可控变量 (A, B, C) 和 4 个噪声变量 (D, E, F, G). 要估计可控变量的主效应和二因子交互效应 (6 个参数)、噪声变量的主效应 (4 个参数), 以及可控变量与噪声变量之间的交互效应 (12 个参数). 包括截距项在内共需估计

23 个参数. 对这类问题, 用 D 最优准则构造的设计是非常好的设计.

表 12.5 是这种情况的含 23 次试验的 D 最优设计. 在这个设计中, 包含可控因子的二因子交互作用不会与其他可控因子的二因子交互作用混杂, 或与其他包含可控因子与噪声因子的二因子交互作用混杂. 然而这些主效应和二因子交互作用与噪声因子的二因子交互作用混杂, 所以, 这个设计的有效性取决于对噪声因子的二因子交互作用是否可忽略的假定.

在拟合可控变量的完全二阶模型时, 中心复合设计 (CCD) 是选择实验设计的合理依据. CCD 可以像例 12.2 那样修改为只有可控变量方向才有轴向点. 例如, 如果有 3 个可控变量和 4 个噪声变量, 在表 12.4 的有 30 个试验点的设计中, 添加关于因子 A, B, C 的 6 个轴向试验点, 以及 4 个中心点, 所得的设计对拟合响应模型而言是非常好的. 所得设计有 40 个试验点, 响应模型有 26 个参数.

表 12.5　3 个可控变量和 4 个噪声变量的含 23 次试验的 D 最优设计

A	B	C	D	E	F	G
−	+	−	+	+	+	+
+	−	−	+	+	+	+
−	−	+	−	−	+	+
+	+	−	−	−	−	+
−	+	+	−	−	−	+
+	+	−	+	−	+	−
+	+	+	−	−	−	−
+	−	−	−	−	+	−
+	−	+	−	−	−	+
−	−	+	−	−	−	−
−	+	+	−	+	+	−
−	−	+	+	+	−	+
−	−	−	−	+	−	−
−	+	−	+	+	−	−
−	−	−	+	−	+	−
+	−	+	+	−	+	−
−	+	+	+	−	+	−
−	−	−	−	−	−	+
+	−	+	−	+	+	−
+	−	−	+	+	−	−
+	+	−	−	+	+	+
−	+	−	−	−	+	−
+	+	+	+	+	−	+

表 12.6　3 个控制变量和 2 个噪声变量的拟合二阶响应模型的 D 最优设计

A	B	C	D	E
+	+	+	+	−
+	+	−	+	+
+	−	−	+	−
0	−	+	−	−
+	+	+	−	+
−	−	−	+	+
−	+	−	+	−

(续)

A	B	C	D	E
−	−	+	+	−
+	+	−	−	−
+	−	−	−	+
+	−	0	−	−
−	−	−	−	−
−	+	−	−	+
−	−	+	−	+
+	0	+	−	−
0	0	0	0	+
−	+	+	+	+
−	+	+	−	−

也可以用其他方法构造二阶情况下的设计. 例如, 假设有 3 个可控因子和 2 个噪声因子. 一个修改过的 CCD 在立方体中有 16 次试验 (2^{5-1}), 在可控变量方向有 6 个轴向试验点, 以及 4 个中心试验点. 这产生了一个有 26 次试验的设计, 可估计有 18 个参数的模型. 另一种方法是用立方体小复合设计 (11 次试验), 并加上在可控变量方向上的 6 个轴向试验点和 4 个中心点. 这个设计总共只有 21 次试验. 也可以用 D 最优方法. 表 12.6 中的有 18 次试验的设计是用 Design-Expert 软件构造的. 这是一个饱和设计. 一般来说, 设计越小, 响应模型中参数就越不能像大的设计那样估计得更好, 预测的方差也越大. RPD 设计和过程稳健性研究方面更多的内容可参阅 Myers and Montgomery (2002) 以及其中的参考文献.

12.6 思 考 题

12.1 再考虑表 12.1 中的簧片实验. 设实验目标是找出一组运行条件, 使得簧片的平均自由高度尽可能接近 7.6 英寸, 同时簧片自由高度的方差尽可能小. 为达到上述目标, 你推荐什么运行条件?

*12.2 考虑思考题 6.18 中的瓶装饮料实验. 设碳酸百分比 (A) 是噪声变量 ($\sigma_z^2 = 1$ 规范单位).

(a) 对这些数据拟合响应模型. 有稳健设计问题吗?

(b) 求出均值模型以及方差模型或 POE.

(c) 找出可以使灌注偏差尽可能接近零并最小化传递方差的一组运行条件.

12.3 考虑思考题 11.12 中的实验. 设温度是噪声变量 ($\sigma_z^2 = 1$ 规范单位). 对两个响应分别拟合响应模型. 关于这两个响应, 有稳健设计问题吗? 找出一组运行条件, 使转化率最大化, 活性在 55 至 60 之间, 并且使温度传递的变异性最小化.

12.4 再考虑表 12.1 中的簧片实验. 设因子 A, B, C 是可控变量, 因子 D 和 E 是噪声因子. 建立一个直积表设计来研究这个问题, 假定所有可控变量的二因子交互作用可忽略不计. 你获得的是哪一种设计?

12.5 **续思考题 12.4**. 再考虑表 12.1 中的簧片实验. 设因子 A, B, C 是可控变量, 因子 D 和 E 是噪声因子. 怎样列出一个组合表设计, 使之可用于估计所有二因子交互作用且只需 16 次试验. 将此设计与思考题 12.4 中的直积表设计进行比较. 从中可以看到, 一般情况下组合表设计比直积表设计的试验次数会如何减少?

12.6 考虑表 12.2 中的连接器拉开力实验. 用这个设计可以估计哪些主效应及可控因子的交互效应? 注意所有可控变量都是定量因子.

12.7 考虑表 12.2 中的连接器拉开力实验. 对这个问题怎样列出一个设计的实验, 使它可拟合这样的模型, 其中包含可控变量的全二次模型的所有项、噪声变量的主效应及噪声变量与可控变量的交互作用. 这个设计需要多少次试验? 比较这个设计与表 12.2 中的设计.

*12.8 考虑思考题 11.11 中的实验. 设压强是噪声变量 ($\sigma_z^2 = 1$ 规范单位). 拟合黏度响应的响应模型. 找出黏度尽可能接近 600 并最小化噪声变量压强所传递变异的一组运行条件.

12.9 例 12.1 的变形. 在例 12.1 (它用例 6.2 的数据) 中, 我们发现过程变量之一 (B = 压强) 并不重要. 删去这个变量后, 就变成一个二次重复的 2^3 设计. 数据如下:

C	D	$A(+)$	$A(-)$	$\bar{y}$	s^2
−	−	45,48	71,65	57.25	161.58
+	−	68,80	60,65	68.25	72.25
−	+	43,45	100,104	73.00	1 124.67
+	+	75,70	86,96	81.75	134.92

假定 C 和 D 是可控因子, A 是噪声因子.

(a) 拟合均值响应的模型.

(b) 拟合 $\ln(s^2)$ 响应的模型.

(c) 找出平均渗透率超过 75 且最小化方差的运行条件.

(d) 将你的结果与例 12.1 中的结果比较, 哪个用了误差传递方法. 两个答案如何类似?

*12.10 在一篇文章中 ("Let's All Beware the Latin Square," *Quality Engineering*, Vol.1, 1989, pp. 453-465), J. S. Hunter 说明了与 3^{k-p} 分式析因设计有关的一些问题. 因子 A 是添加到标准燃料中乙醇的量, B 表示空气/燃料比. 响应变量是一氧化碳 (CO) 排放量, 单位 g/m^3. 设计如下.

设计				观测	
A	B	x_1	x_2	y	
0	0	−1	−1	66	62
1	0	0	−1	78	81
2	0	+1	−1	90	94
0	1	−1	0	72	67
1	1	0	0	80	81
2	1	+1	0	75	78
0	2	−1	+1	68	66
1	2	0	+1	66	69
2	2	+1	+1	60	58

注意到我们已经使用了 0, 1, 2 符号系统来表示因子的低, 中, 高水平. 我们也使用了 −1, 0, +1 的 "几何符号" 系统. 设计中的每个试验重复两次.

(a) 验证二阶模型

$$\hat{y} = 78.5 + 4.5x_1 - 7.0x_2 - 4.5x_1^2 - 4.0x_2^2 - 9.0x_1x_2$$

是该实验的一个合理的模型. 画出 CO 浓度在 x_1, x_2 空间上等高线的草图.

(b) 假定用四因子的 3^{4-2} 分式析因设计取代二因子获得与 (a) 中完全相同的数据. 设计将会是:

设计								观测	
A	B	C	D	x_1	x_2	x_3	x_4	y	
0	0	0	0	−1	−1	−1	−1	66	62
1	0	1	1	0	−1	0	0	78	81
2	0	2	2	+1	−1	+1	+1	90	94
0	1	2	1	−1	0	+1	0	72	67
1	1	0	2	0	0	−1	+1	80	81
2	1	1	0	+1	0	0	−1	75	78
0	2	1	2	−1	+1	0	+1	68	66
1	2	2	0	0	+1	+1	−1	66	69
2	2	0	1	+1	+1	−1	0	60	58

计算 CO 响应在 4 个因子 A, B, C, D 每个水平上的边际平均. 画出这些边际平均的图形并解释结果. 可以看出因子 C 和 D 有强的效应吗? 这些因子的确对 CO 排放有影响吗? 为什么它们有强的效应?

(c) 用 (b) 中的设计可拟合模型

$$y = \beta_0 + \sum_{i=1}^{4} \beta_i x_i + \sum_{i=1}^{4} \beta_i x_i^2 + \varepsilon$$

假定真正的模型是

$$y = \beta_0 + \sum_{i=1}^{4} \beta_i x_i + \sum_{i=1}^{4} \beta_{ij} x_i^2 + \sum_{i<j}\sum \beta_{ij} x_i x_j + \varepsilon$$

证明, 如果 $\hat{\beta}_i$ 表示拟合模型中系数的最小二乘估计, 则

$$\begin{aligned}
&\mathrm{E}(\hat{\beta}_0) = \beta_0 - \beta_{13} - \beta_{14} - \beta_{34} && \mathrm{E}(\hat{\beta}_{11}) = \beta_{11} - (\beta_{23} - \beta_{24})/2\\
&\mathrm{E}(\hat{\beta}_1) = \beta_1 - (\beta_{23} + \beta_{24})/2 && \mathrm{E}(\hat{\beta}_{22}) = \beta_{22} + (\beta_{13} + \beta_{14} + \beta_{34})/2\\
&\mathrm{E}(\hat{\beta}_2) = \beta_2 - (\beta_{13} + \beta_{14} + \beta_{34})/2 && \mathrm{E}(\hat{\beta}_{33}) = \beta_{33} - (\beta_{24} - \beta_{12})/2 + \beta_{14}\\
&\mathrm{E}(\hat{\beta}_3) = \beta_3 - (\beta_{12} + \beta_{24})/2 && \mathrm{E}(\hat{\beta}_{44}) = \beta_{44} - (\beta_{12} - \beta_{23})/2 + \beta_{13}\\
&\mathrm{E}(\hat{\beta}_4) = \beta_4 - (\beta_{12} + \beta_{23})/2
\end{aligned}$$

这有助于解释在 (b) 中所观测到的因子 C 和 D 的强效应吗?

12.11 在晶片涂层的工序中进行实验. 实验中的每次试验生产一个晶片, 并在晶片的不同位置测量若干个涂层的厚度. 得到涂层厚度的平均值 y_1 和标准差 y_2. 数据 [根据 Box and Draper (1987) 改写] 如下表:

试验号	速度	压强	距离	均值 y_1	标准差 y_2	试验号	速度	压强	距离	均值 y_1	标准差 y_2
1	−1	−1	−1	24.0	12.5	15	+1	0	0	501.7	92.5
2	0	−1	−1	120.3	8.4	16	−1	+1	0	264.0	63.5
3	+1	−1	−1	213.7	42.8	17	0	+1	0	427.0	88.6
4	−1	0	−1	86.0	3.5	18	+1	+1	0	730.7	21.1
5	0	0	−1	136.6	80.4	19	−1	−1	+1	220.7	133.8
6	+1	0	−1	340.7	16.2	20	0	−1	+1	239.7	23.5
7	−1	+1	−1	112.3	27.6	21	+1	−1	+1	422.0	18.5
8	0	+1	−1	256.3	4.6	22	−1	0	+1	199.0	29.4
9	+1	+1	−1	271.7	23.6	23	0	0	+1	485.3	44.7
10	−1	−1	0	81.0	0.0	24	+1	0	+1	673.7	158.2
11	0	−1	0	101.7	17.7	25	−1	+1	+1	176.7	55.5
12	+1	−1	0	357.0	32.9	26	0	+1	+1	501.0	138.9
13	−1	0	0	171.3	15.0	27	+1	+1	+1	1 010.0	142.4
14	0	0	0	372.0	0.0						

(a) 实验所用的是哪种类型的设计? 它是拟合二次模型好的选择吗?

(b) 建立两个响应的模型.

(c) 找出均值尽可能大且标准差小于 60 的一组最优运行条件.

*12.12 假设有 4 个可控变量和 2 个噪声变量. 需要估计所有可控变量的主效应和二因子交互作用、噪声变量的主效应以及所有可控因子与噪声因子之间的二因子交互作用. 若所有因子有两个水平, 可估计所有模型参数的组合表设计最小试验次数是多少? 用 D 最优算法找出一个设计.

*12.13 设有 4 个可控变量和 2 个噪声变量. 需要拟合的模型包括: 可控变量的完全二次模型的所有项、噪声变量的主效应, 以及所有可控因子与噪声因子之间的二因子交互作用. 用修改中心复合设计的方式对此问题建立一个组合表设计.

12.14 再考虑思考题 12.13 的情况. 对此问题可以使用改进的小复合设计吗? 用小复合设计有什么缺点?

12.15 再考虑思考题 12.13 的情况. 可估计所有模型参数的组合表设计的最少试验次数是多少? 用 D 最优算法对此问题找出一个合理的设计.

12.16 对波动焊接过程做一个实验. 有 5 个可控变量和 3 个噪声变量. 响应变量是每百万次焊接中焊接缺陷的个数. 所用的实验设计是如下所示的直积表.

内表					外表				
					F	−1	1	1	−1
A	B	C	D	E	G	−1	1	−1	1
					H	−1	−1	1	1
1	1	1	−1	−1		194	197	193	275
1	1	−1	1	1		136	136	132	136
1	−1	1	−1	1		185	261	264	264
1	−1	−1	1	−1		47	125	127	42
−1	1	1	1	−1		295	216	204	293
−1	1	−1	−1	1		234	159	231	157
−1	−1	1	1	1		328	326	247	322
−1	−1	−1	−1	−1		186	187	105	104

(a) 内表和外表所用的分别是哪种类型的设计? 这些设计的别名关系是什么?

(b) 建立焊接缺陷的均值模型与方差模型. 你会推荐什么操作条件?

12.17 再考虑思考题 12.16 中的波动焊接实验. 找出此实验需要较少试验次数的组合表设计.

12.18 再考虑思考题 12.16 中的波动焊接实验. 设需要拟合的模型包括: 可控变量的完全二次模型的所有项、噪声变量的所有主效应, 以及所有可控变量与噪声变量之间的交互作用. 你推荐什么设计?

第 13 章　含随机因子的实验

本章纲要

13.1　随机效应模型
13.2　含随机因子的二因子析因设计
13.3　二因子混合模型
13.4　含随机效应的样本量的确定
13.5　期望均方的计算法则
13.6　近似 F 检验
13.7　关于方差分量估计的一些其他论题
13.7.1　方差分量的近似置信区间
13.7.2　修正的大样本方法
13.7.3　方差分量的最大似然估计
第 13 章补充材料
S13.1　随机模型的期望均方
S13.2　混合模型的期望均方
S13.3　约束混合模型与无约束混合模型
S13.4　样本容量不等的随机与混合模型
S13.5　关于修正的大样本方法的背景材料
S13.6　基于修正的大样本方法的方差分量比的置信区间

本书大部分都假定实验中的因子是**固定因子**, 即实验者采用的因子水平是感兴趣的特定水平. 这意味着, 对因子所做的统计推断被限制于所研究的特定水平. 在例 5.1 电池寿命实验中研究 3 种材料, 我们的结论只对这些特定的材料类型有效. 当一个或多个因子是**定量因子**时, 会出现与此不同的情况. 这时, 我们通常用一个把响应与因子相联系的回归模型来预测实验设计中因子水平范围内的响应. 第 5 章到第 9 章中已给出几个例子. 一般来讲, 对于固定效应, 实验的**推断空间**是指所研究的一组特定的因子水平.

在某些实验中, 因子水平是从大量的可能水平中随机选出的, 实验者想从中得出适用于所有水平的结论, 而不仅仅是适用于那些实验设计中用到的水平. 此时, 这种因子称为**随机因子**. 我们先从一个简单的情况开始, 单因子实验中的因子是随机的, 以此引进用于方差分析和**方差分量**分析的**随机效应模型**. 随机因子也经常在析因实验以及其他类型的实验中出现. 本章主要讨论有随机因子的析因实验的设计与分析的方法. 第 14 章将介绍的**嵌套设计**与**裂区设计**在实际中经常会遇到随机因子.

13.1　随机效应模型

实验者经常会对某个有许多可能水平的因子感兴趣. 若实验者从这众多水平中随机选取 a 个水平, 则称该因子是**随机的**. 由于实验中采用的该因子的水平是随机选取的, 相应的推断对因子的全体水平都有效. 假定因子的水平总数是无限的或大到可以认为是无限的. 通常不会遇到随机因子的水平总数小到只能利用有限总体方法的情况. 关于有限总体情况时的讨论, 可以参见 Bennett and Franklin(1954) 以及 Searle and Fawcett (1970).

线性统计模型为

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, n \end{cases} \tag{13.1}$$

其中 τ_i 与 ε_{ij} 都是随机变量. 若 τ_i 的方差为 σ_τ^2, 且与 ε_{ij} 独立, 则任一观测值的方差为

$$\mathrm{V}(y_{ij}) = \sigma_\tau^2 + \sigma^2$$

方差 σ_τ^2 与 σ^2 称为**方差分量**, 而模型 [(13.1) 式] 称为**方差分量或随机效应模型**. 为了检验该模型的假设, 我们要求 $\{\varepsilon_{ij}\}$ 是 NID$(0,\sigma^2)$, $\{\tau_i\}$ 是 NID$(0,\sigma_\tau^2)$, 且 τ_i 与 ε_{ij} 相互独立①.

方差分析的平方和分解式

$$SS_T = SS_{处理} + SS_E \tag{13.2}$$

仍然有效. 即我们将观测中的总差异分解成两部分, 一部分度量了处理间的差异 ($SS_{处理}$), 另一部分度量了处理内部的差异 (SS_E). 检验单个处理效应是没有意义的, 所以我们对方差分量 σ_τ^2 的假设进行检验:

$$\begin{aligned} H_0 &: \sigma_\tau^2 = 0 \\ H_1 &: \sigma_\tau^2 > 0 \end{aligned} \tag{13.3}$$

若 $\sigma_\tau^2 = 0$, 则所有处理都相等; 若 $\sigma_\tau^2 > 0$, 则处理间存在差异. 和以前一样, SS_E/σ^2 服从自由度为 $N-a$ 的卡方分布, 在零假设下, $SS_{处理}/\sigma^2$ 服从自由度为 $a-1$ 的卡方分布. 所有随机变量都相互独立. 这样, 在零假设下, 比值

$$F_0 = \frac{\frac{SS_{处理}}{a-1}}{\frac{SS_E}{N-a}} = \frac{MS_{处理}}{MS_E} \tag{13.4}$$

服从自由度为 $a-1$ 和 $N-a$ 的 F 分布. 但是, 为了说明检验方法, 我们先计算期望均方值.

考虑

$$\begin{aligned} \mathrm{E}(MS_{处理}) &= \frac{1}{a-1}\mathrm{E}(SS_{处理}) = \frac{1}{a-1}\mathrm{E}\left[\sum_{i=1}^{a}\frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}\right] \\ &= \frac{1}{a-1}\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{a}\left(\sum_{j=1}^{n}\mu+\tau_i+\varepsilon_{ij}\right)^2 - \frac{1}{N}\left(\sum_{i=1}^{a}\sum_{j=1}^{n}\mu+\tau_i+\varepsilon_{ij}\right)^2\right] \end{aligned}$$

因为 $\mathrm{E}(\tau_i) = 0$, 所以平方后逐项求期望时, 包含 τ_i^2 的项都用 σ_τ^2 替换. 同时, 包含 $\varepsilon_{i.}^2$、$\varepsilon_{..}^2$ 和 $\sum_{i=1}^{a}\sum_{j=1}^{n}\tau_i^2$ 的项分别用 $n\sigma^2$、$an\sigma^2$ 和 $an^2\sigma_\tau^2$ 替换. 而且, 所有包含 τ_i 或 ε_{ij} 的交叉项的期望都为零. 所以

$$\mathrm{E}(MS_{处理}) = \frac{1}{a-1}\left[N\mu^2 + N\sigma_\tau^2 + a\sigma^2 - N\mu^2 - n\sigma_\tau^2 - \sigma^2\right]$$

即

$$\mathrm{E}(MS_{处理}) = \sigma^2 + n\sigma_\tau^2 \tag{13.5}$$

类似地, 可以证明

$$\mathrm{E}(MS_E) = \sigma^2 \tag{13.6}$$

由期望均方可知, 在 H_0 下检验统计量 [(13.4) 式] 的分子和分母都是 σ^2 的无偏估计, 而在 H_1 下分子的期望大于分母的期望. 所以 F_0 的值特别大时我们应该拒绝 H_0. 这是上尾部的单边拒绝域, 因此, 当 $F_0 > F_{\alpha,a-1,N-a}$ 时我们拒绝 H_0.

---

① $\{\tau_i\}$ 是独立的随机变量这一假设意味着, 固定效应模型中的一般假设 $\sum_{i=1}^{a}\tau_i = 0$ 并不应用于随机效应模型.

随机效应模型的计算步骤和方差分析类似于固定效应的情形. 不过结论对处理的所有水平有效, 这点上有很大不同.

我们通常需要估计模型中的方差分量 (σ_τ^2 与 σ^2). 用于估计 σ_τ^2 与 σ^2 的方法称为**方差分析法**, 因为它利用方差分析表得到. 计算步骤为: 先令方差分析表中的各期望均方等于它们的观测值, 然后解得方差分量. 在单因子随机效应模型中, 令各均方观测值等于它们的期望值, 得到

$$MS_{\text{处理}} = \sigma^2 + n\sigma_\tau^2, \qquad MS_E = \sigma^2$$

所以, 方差分量的估计为

$$\hat{\sigma}^2 = MS_E, \tag{13.7}$$

$$\hat{\sigma}_\tau^2 = \frac{MS_{\text{处理}} - MS_E}{n} \tag{13.8}$$

样本容量不等时, 把 (13.8) 式中的 n 换成

$$n_0 = \frac{1}{a-1}\left[\sum_{i=1}^{a} n_i - \frac{\sum_{i=1}^{a} n_i^2}{\sum_{i=1}^{a} n_i}\right] \tag{13.9}$$

估计方差分量的方差分析法并不要求正态假定. 得到的 σ_τ^2 与 σ^2 估计是最优二次无偏的(即在由观测值的二次函数构成的所有无偏估计中, 这类估计量的方差最小).

有时方差分析法会得到负的方差分量估计值. 显然, 由定义知, 方差分量是非负的, 所以负的方差分量估计值应该被特别对待. 一种做法是, 假定抽样波动导致负的估计值, 则可接受该估计, 并以此作为方差分量的实际值为零的依据. 这种做法很直观, 但在理论上有一些问题. 例如, 用零代替负的估计值会影响其他估计的统计性质. 另一种做法是, 采用总是得到非负估计的其他方法重新估计这个负的方差分量. 还有一种做法是, 得到负估计说明线性模型的假定是错误的, 应该重新检查线性假定. 方差分量估计的更多内容参见 Searle(1971a, 1971b), Searle, Casella, and McCullogh(1992), 以及 Burdick and Graybill(1990).

例 13.1 一家纺织厂用许多台织机编织一种织物, 希望织机尽量相同使得织物的强度一致. 过程工程师怀疑, 除了同一台织机编织的织物间强度会有差异外, 在不同的织机间织物的强度也有显著差异. 为此, 她随机选择了 4 台织机, 并对每台织机生产的织物进行了 4 次强度测量. 实验按随机顺序进行, 数据如表 13.1 所示. 进行方差分析, 其结果列在表 13.2 上. 根据方差分析, 得出工厂的织机间存在显著差异的结论.

表 13.1 例 13.1 的强度数据

织机	观测值				$y_{i\cdot}$
	1	2	3	4	
1	98	97	99	96	390
2	91	90	93	92	366
3	96	95	97	95	383
4	95	96	99	98	388
					$1\ 527 = y_{\cdot\cdot}$

方差分量的估计为 $\hat{\sigma}^2 = 1.90$, 且

$$\hat{\sigma}_\tau^2 = \frac{29.73 - 1.90}{4} = 6.96$$

表 13.2 例 13.1 强度数据的方差分析

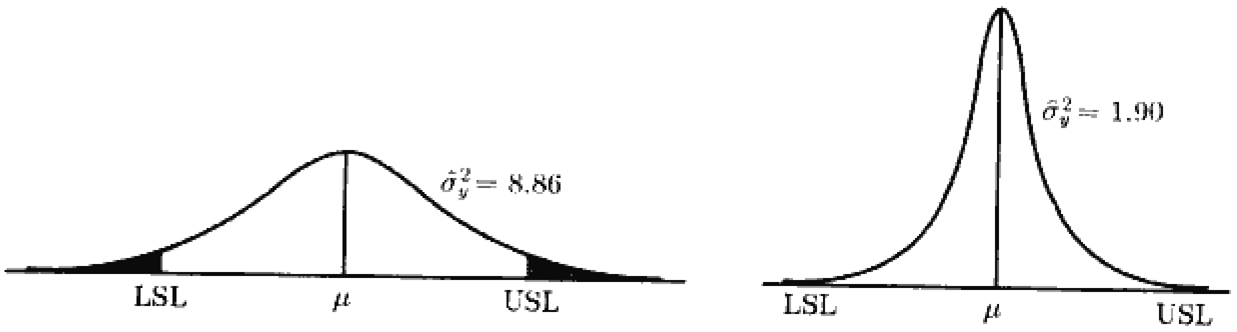
方差来源	平方和	自由度	均方	F_0	P 值
织机	89.19	3	29.73	15.68	<0.001
误差	22.75	12	1.90		
和	111.94	15			

所以, 任一强度观测的方差的估计为

$$\hat{\sigma}_y = \hat{\sigma}^2 + \hat{\sigma}_\tau^2 = 1.90 + 6.96 = 8.86$$

这里大部分的变异源于织机间的差异.

上述例子说明方差分量的一个重要应用 —— 分离影响一个产品或系统的不同变异来源. 质量保证中经常会遇到产品的变异性问题, 但要分离变异的来源通常比较困难. 例如, 该研究也许缘起于观测到织物的强度间有很大的变异性, 如图 13.1a 所示. 该图显示了将过程产出 (织物强度) 看作方差 $\hat{\sigma}_y^2 = 8.86$(即例 13.1 中对强度观测值的方差的估计值) 的正态分布. 强度规格的上限与下限也都标示在图 13.1a 上, 容易看出, 过程产出中有相当大的比例落在规格限之外 (图 13.1a 的尾部阴影区域). 过程工程师疑虑为什么这么多的织物不合格, 需要报废、返修或降格为低质量产品. 答案是产品强度的差异主要源于织机间的差异. 织机间性能的差异可能源于错误的设置、缺乏保养、缺少监管、操作员技术差、原材料有问题等.



(a) 过程产出的变异　　(b) 当 $\sigma_\tau^2 = 0$ 时的过程产出的变异

图 13.1 纤维性织物强度问题中的过程产出

过程工程师现在需要区分造成织机性能差异的具体原因. 若她能找出并消除织机间差异的根源, 过程产出的方差就可以明显减少, 也许可以低到 $\hat{\sigma}_y^2 = 1.90$, 即例 13.1 中织机内 (随机误差) 方差分量的估计值. 图 13.1b 显示的是织物强度的方差为 $\hat{\sigma}_y^2 = 1.90$ 的正态分布, 注意此时产出中的不合格品率已被大大降低. 虽然不太可能根除织机间的所有差异, 但是, 方差分量的明显减小显然能大大提高所产织物的质量.

容易求得方差分量 σ^2 的置信区间. 若观测值是正态的且相互独立, 则 $(N-a)MS_E/\sigma^2$ 的分布为 χ^2_{N-a}. 因此有

$$\mathrm{P}\left[\chi^2_{1-(\alpha/2),N-a} \leqslant \frac{(N-a)MS_E}{\sigma^2} \leqslant \chi^2_{\alpha/2,N-a}\right] = 1-\alpha$$

从而 σ^2 的 $100(1-\alpha)\%$ 的置信区间为

$$\frac{(N-a)MS_E}{\chi^2_{\alpha/2,N-a}} \leqslant \sigma^2 \leqslant \frac{(N-a)MS_E}{\chi^2_{1-(\alpha/2),N-a}} \tag{13.10}$$

现在考虑方差分量 σ_τ^2 的置信区间. σ_τ^2 的点估计为

$$\hat{\sigma}_\tau^2 = \frac{MS_{处理} - MS_E}{n}$$

随机变量 $(a-1)MS_{处理}/(\sigma^2 + n\sigma_\tau^2)$ 的分布为 χ^2_{a-1}, $(N-a)MS_E/\sigma^2$ 的分布为 χ^2_{N-a}. 所以 $\hat{\sigma}_\tau^2$ 的概率分布为两个卡方分布随机变量的线性组合, 即

$$u_1\chi^2_{a-1} - u_2\chi^2_{N-a}$$

其中

$$u_1 = \frac{\sigma^2 + n\sigma_\tau^2}{n(a-1)}, \quad u_2 = \frac{\sigma^2}{n(N-a)}$$

不幸的是, 对于这个卡方分布随机变量的线性组合的分布而言无法得到其精确表达式. 这样, 就不能构造 σ_τ^2 的精确置信区间. 近似算法参见 Graybill(1961) 和 Searle(1971a). 也可以参见 13.7 节.

容易得到关于比值 $\sigma_\tau^2/(\sigma^2 + \sigma_\tau^2)$ 的置信区间的精确表达式. 该比值称为**组内相关系数**, 它反映了, 在观测值的方差中, 由处理间的差异产生的那部分所占的比例. 为了推导在平衡设计情形下它的置信区间, 注意到 MS_E 与 $MS_{处理}$ 是相互独立的随机变量, 而且可以证明

$$\frac{MS_{处理}/(n\sigma_\tau^2 + \sigma^2)}{MS_E/\sigma^2} \sim F_{a-1,N-a}$$

所以,

$$(F_{1-\alpha/2,a-1,N-a} \leqslant \frac{MS_{处理}}{MS_E}\frac{\sigma^2}{n\sigma_\tau^2 + \sigma^2} \leqslant F_{\alpha/2,a-1,N-a}) = 1 - \alpha \tag{13.11}$$

将 (13.11) 式变形, 可以得到:

$$\mathrm{P}\left(L \leqslant \frac{\sigma_\tau^2}{\sigma^2} \leqslant U\right) = 1 - \alpha \tag{13.12}$$

其中

$$L = \frac{1}{n}\left(\frac{MS_{处理}}{MS_E}\frac{1}{F_{\alpha/2,a-1,N-a}} - 1\right) \tag{13.13a}$$

$$U = \frac{1}{n}\left(\frac{MS_{处理}}{MS_E}\frac{1}{F_{1-\alpha/2,a-1,N-a}} - 1\right) \tag{13.13b}$$

这里 L 与 U 分别是比值 σ_τ^2/σ^2 的 $100(1-\alpha)\%$ 置信下限与上限. 所以比值 $\sigma_\tau^2/(\sigma^2 + \sigma_\tau^2)$ 的 $100(1-\alpha)\%$ 置信区间为

$$\frac{L}{1+L} \leqslant \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma^2} \leqslant \frac{U}{1+U} \tag{13.14}$$

为说明该方法, 我们根据例 13.1 中的强度数据计算 $\sigma_\tau^2/(\sigma^2+\sigma_\tau^2)$ 的 95% 的置信区间. 由于 $MS_{处理} = 29.73$, $MS_E = 1.90$, $a = 4$, $n = 4$, $F_{0.025,3,12} = 4.47$, 且 $F_{0.975,3,12} = \frac{1}{F_{0.025,12,3}} = \frac{1}{14.34} = 0.070$. 所以由等式 (13.13a) 和 (13.13b),

$$L = \frac{1}{4}\left(\frac{29.73}{1.90} \times \frac{1}{4.47} - 1\right) = 0.625, \qquad U = \frac{1}{4}\left(\frac{29.73}{1.90} \times \frac{1}{0.070} - 1\right) = 56.633$$

由 (13.14) 式, $\sigma_\tau^2/(\sigma^2+\sigma_\tau^2)$ 的 95% 的置信区间为

$$\frac{0.625}{1.625} \leqslant \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \leqslant \frac{55.633}{56.633}$$

即

$$0.38 \leqslant \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \leqslant 0.98$$

我们认为在织物强度观测的方差中织机间的变异占了 38%到 98%的比重. 这里的置信区间比较宽是因为实验中的样本容量较小. 不过, 显然织机间的变异是不能忽略的.

13.2 含随机因子的二因子析因设计

假定有两个因子 A 和 B, 它们都有许多水平受到关注 (如上节所述, 假定水平的总数是无穷多). 随机选取因子 A 的 a 个水平, 因子 B 的 b 个水平, 并采用析因实验设计来安排它们的水平组合. 若实验重复 n 次, 我们可用线性模型

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \quad \begin{cases} i = 1,2,\cdots,a \\ j = 1,2,\cdots,b \\ k = 1,2,\cdots,n \end{cases} \tag{13.15}$$

描述观测值, 其中模型参数 $\tau_i, \beta_j, (\tau\beta)_{ij}, \varepsilon_{ijk}$ 是随机变量. 还要假定随机变量 $\tau_i, \beta_j, (\tau\beta)_{ij}, \varepsilon_{ijk}$ 都服从均值为 0 的正态分布, 方差分别为 $\mathrm{V}(\tau_i)=\sigma_\tau^2$, $\mathrm{V}(\beta_j)=\sigma_\beta^2$, $\mathrm{V}[(\tau\beta)_{ij}]=\sigma_{\tau\beta}^2$, $\mathrm{V}(\varepsilon_{ijk})=\sigma^2$. 任一观测值的方差是

$$\mathrm{V}(y_{ijk}) = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_{\tau\beta}^2 + \sigma^2 \tag{13.16}$$

且 $\sigma_\tau^2, \sigma_\beta^2, \sigma_{\tau\beta}^2, \sigma^2$ 称为**方差分量**. 我们要检验的假设是 $H_0: \sigma_\tau^2 = 0, H_0: \sigma_\beta^2 = 0$, $H_0: \sigma_{\tau\beta}^2 = 0$. 注意它们与单因子随机效应模型的相似性. 方差分析中的数值计算不变. 也就是说, SS_A, SS_B, SS_{AB}, SS_E 的计算方法都和固定效应的情况相同. 但是要给出检验统计量, 须先考察各**期望均方**. 可以证明

$$\mathrm{E}(MS_A) = \sigma^2 + n\sigma_{\tau\beta}^2 + bn\sigma_\tau^2, \qquad \mathrm{E}(MS_B) = \sigma^2 + n\sigma_{\tau\beta}^2 + an\sigma_\beta^2 \tag{13.17}$$

$$\mathrm{E}(MS_{AB}) = \sigma^2 + n\sigma_{\tau\beta}^2, \qquad \mathrm{E}(MS_E) = \sigma^2$$

由期望均方看出, 用来检验无交互作用假设 $H_0: \sigma_{\tau\beta}^2 = 0$ 的恰当的统计量是

$$F_0 = \frac{MS_{AB}}{MS_E} \tag{13.18}$$

因为在 H_0 下, F_0 的分子与分母的期望都是 σ^2, 而且, 仅当 H_0 不真时, 才有 $\mathrm{E}(MS_{AB})$ 大于 $\mathrm{E}(MS_E)$. 比值 F_0 的分布为 $F_{(a-1)(b-1),ab(n-1)}$. 同理, 要检验 $H_0: \sigma_\tau^2 = 0$, 可用

$$F_0 = \frac{MS_A}{MS_{AB}} \tag{13.19}$$

其分布为 $F_{a-1,(a-1)(b-1)}$; 要检验 $H_0: \sigma_\beta^2 = 0$, 统计量是

$$F_0 = \frac{MS_B}{MS_{AB}} \tag{13.20}$$

其分布为 $F_{b-1,(a-1)(b-1)}$. 这些都是上尾部的单边检验. 注意, 这些检验统计量与因子 A 和 B 都是固定效应时所用的检验统计量不同. 均方的期望常常可以用来指导检验统计量的构造.

在许多涉及随机因子的实验中, 方差分量的估计至少受到与假设检验同等程度的关注. 可用**方差分析法**来求出方差分量的估计. 也就是说, 令方差分析表中观测到的均方值等于它们的期望值, 进而解得方差分量. 得出

$$\hat{\sigma}^2 = MS_E, \qquad \hat{\sigma}^2_{\tau\beta} = \frac{MS_{AB} - MS_E}{n} \tag{13.21}$$

$$\hat{\sigma}^2_{\beta} = \frac{MS_B - MS_{AB}}{an}, \qquad \hat{\sigma}^2_{\tau} = \frac{MS_A - MS_{AB}}{bn}$$

作为二因子随机效应模型中方差分量的点估计. 我们将在 13.7 节中讨论方差分量点估计的其他方法, 以及构造置信区间的方法.

例 13.2 测量系统的能力研究

统计设计的实验经常用于研究影响一个系统的变异来源. 一个常见的工业应用是用实验设计去研究一个测量系统的变异成分. 这类研究常称为**量具能力研究 (gauge capability studies)**或**量具可重复性与可再现性 (gauge repeatability and reproducibility, R&R) 研究**, 因为它们是受到关注的变异成分 (关于 R&R 研究的更多讨论, 参见本章补充材料).

一个典型的 R&R 实验 [来自 Montgomery(2001)] 见表 13.3. 量具用于测量工件的关键尺寸. 从生产线上选出 20 个工件, 随机选择 3 名操作员用这种量具测量每个工件两次. 测量次序是完全随机的, 所以这是一个二因子析因实验, 设计因子是工件与操作员, 有两次重复. 工件与操作员都是随机因子. 应用 (13.16) 式中的方差分量等式, 即

$$\sigma^2_y = \sigma^2_\tau + \sigma^2_\beta + \sigma^2_{\tau\beta} + \sigma^2$$

表 13.3 例 13.2 的测量系统能力实验

工件编号	操作员 1		操作员 2		操作员 3	
1	21	20	20	20	19	21
2	24	23	24	24	23	24
3	20	21	19	21	20	22
4	27	27	28	26	27	28
5	19	18	19	18	18	21
6	23	21	24	21	23	22
7	22	21	22	24	22	20
8	19	17	18	20	19	18
9	24	23	25	23	24	24
10	25	23	26	25	24	25
11	21	20	20	20	21	20
12	18	19	17	19	18	19
13	23	25	25	25	25	25
14	24	24	23	25	24	25
15	29	30	30	28	31	30
16	26	26	25	26	25	27
17	20	20	19	20	20	20
18	19	21	19	19	21	23
19	25	26	25	24	25	25
20	19	19	18	17	19	17

其中 σ_y^2 是总变异 (包括了由不同工件、操作员以及量具造成的变异), σ_τ^2 是工件的方差分量, σ_β^2 是操作员的方差分量, $\sigma_{\tau\beta}^2$ 是工件与操作员的交互作用的方差分量, σ^2 是随机实验误差的方差. 特别地, 由于 σ^2 被认为是反映了由同一名操作员测量同一工件时观测到的变异, 所以方差分量 σ^2 称为量具的可重复性. 而 $\sigma_\beta^2+\sigma_{\tau\beta}^2$ 反映了操作员使用量具所造成的测量系统中的其余的变异, 所以它被称为量具的可再现性.

表 13.4 列出了本实验的方差分析. 计算是由 Minitab 中的 Balanced ANOVA 过程完成的. 根据 P 值, 我们认为工件的效应是大的, 操作员也许稍微有些影响, 工件 — 操作员交互作用没有显著效应. 用 13.21 式估计各方差分量如下:

$$\hat{\sigma}_\tau^2=\frac{62.39-0.71}{(3)(2)}=10.28,\quad \hat{\sigma}_{\tau\beta}^2=\frac{0.71-0.99}{2}=-0.14$$

$$\hat{\sigma}_\beta^2=\frac{1.31-0.71}{(20)(2)}=0.015,\quad \hat{\sigma}^2=0.99$$

表 13.4　例 13.2 的方差分析 (Minitab Balanced ANOVA)

```
Analysis of Variance (Balanced Designs)

Factor     Type Levels Values
part      random    20     1     2     3     4     5     6     7
                           8     9    10    11    12    13    14
                          15    16    17    18    19    20
operator random      3     1     2     3

Analysis of Variance for y

Source            DF        SS        MS       F       P
part              19   1185.425    62.391   87.65   0.000
operator           2      2.617     1.308    1.84   0.173
part*operator     38     27.050     0.712    0.72   0.861
Error             60     59.500     0.992
Total            119   1274.592

Source            Variance   Error  Expected Mean Square for Each Term
                  component  term   (using unrestricted model)
1 part             10.2798     3    (4) + 2(3) +  6(1)
2 operator          0.0149     3    (4) + 2(3) + 40(2)
3 part*operator    -0.1399     4    (4) + 2(3)
4 Error             0.9917          (4)
```

Minitab 输出的表 13.4 的底部包含了随机模型的各期望均方, 带圆括号的数字代表了方差分量 [(4) 代表 σ^2, (3) 代表 $\sigma_{\tau\beta}^2$, 等等]. 还给出了方差分量的估计, 以及用于检验方差分析中的方差分量的误差项. 术语 "unrestricted model(无约束模型)" 与随机模型无关, 它要在后面才讨论.

注意到有个方差分量 (即 $\sigma_{\tau\beta}^2$) 的估计是负的. 这当然不合理, 因为根据定义方差是非负的. 遗憾的是, 采用方差分析法进行估计, 可能会得到负的方差分量估计值. 可以用不同的方法来处理这种负的估计值. 一种方法是, 假定负的估计值意味着方差分量实际值为零, 于是就设其为零, 其余方差分量估计值不变. 另一种方法是, 用一种确保估计值非负的方法来估计方差分量. 最后, 注意到表 13.4 中交互作用项的 P 值很大,

这可看作 $\sigma^2_{\tau\beta}$ 实际值为零且没有交互效应的证据, 因此拟合如下的**简化模型**:

$$y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk}$$

其中不含交互作用项. 这种方法相对简单, 很多时候与更复杂的方法几乎同样有效.

表 13.5 显示了简化模型的方差分析. 由于模型中没有交互作用项, 两个主效应都与误差项相比来进行检验, 各方差分量的估计如下:

$$\hat{\sigma}^2_\tau = \frac{62.39 - 0.88}{(3)(2)} = 10.25, \qquad \hat{\sigma}^2_\beta = \frac{1.31 - 0.88}{(20)(2)} = 0.010\ 8, \qquad \hat{\sigma}^2 = 0.88$$

表 13.5 例 13.2 的简化模型的方差分析

```
Analysis of Variance (Balanced Designs)

Factor      Type Levels Values
part      random     20      1     2     3     4     5     6     7
                             8     9    10    11    12    13    14
                            15    16    17    18    19    20
operator random       3      1     2     3

Analysis of Variance for y

Source              DF          SS         MS        F       P
part                19    1185.425     62.391    70.64   0.000
operator             2       2.617      1.308     1.48   0.232
Error               98      86.550      0.883
Total              119    1274.592

Source          Variance   Error   Expected Mean Square for Each Term
               component    term   (using unrestricted model)
1 part           10.2513       3   (3) +  6(1)
2 operator        0.0106       3   (3) + 40(2)
3 Error           0.8832           (3)
```

最后, 用方差分量估计 $\hat{\sigma}^2$ 与 $\hat{\sigma}^2_\beta$ 的和来估计量具的方差:

$$\hat{\sigma}^2_{\text{量规}} = \hat{\sigma}^2 + \hat{\sigma}^2_\beta = 0.88 + 0.010\ 8 = 0.890\ 8$$

量具的变异性相对于产品的变异性显然为小. 这是一种理想的情况, 说明量具能够区分不同等级的产品.

测量系统能力研究是实验设计的常见的应用. 这些实验中几乎都有随机效应. 要了解更多的内容或参考文献, 参见 Burdick, Borror, and Montgomery(2003).

13.3 二因子混合模型

现在考虑因子 A 固定而因子 B 随机的情况, 称之为方差分析**混合模型**. 线性统计模型为

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \quad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b \\ k = 1, 2, \cdots, n \end{cases} \tag{13.22}$$

其中 τ_i 是固定效应, β_j 是随机效应, 假定交互作用 $(\tau\beta)_{ij}$ 是随机效应, 而 ε_{ijk} 是随机误差. 还假定 $\{\tau_i\}$ 是使得 $\sum\limits_{i=1}^{a} \tau_i = 0$ 的固定效应, 而 β_j 是 $\text{NID}(0, \sigma_\beta^2)$ 的随机变量. 交互作用效应 $(\tau\beta)_{ij}$ 是正态随机变量, 其均值为 0, 方差为 $[(a-1)/a]\sigma_{\tau\beta}^2$. 但交互作用分量按固定效应加起来为零. 这就是说

$$\sum_{i=1}^{a} (\tau\beta)_{ij} = (\tau\beta)_{\cdot j} = 0 \qquad j = 1,\ 2, \cdots, b$$

这意味着, 固定因子不同水平上的某些交互作用元素不是独立的. 事实上, 可以证明 (见思考题 13.25)

$$\text{Cov}[(\tau\beta)_{ij}, (\tau\beta)_{i'j}] = -\frac{1}{a}\sigma_{\tau\beta}^2 \qquad i \neq i'$$

而对 $j \neq j'$, $(\tau\beta)_{ij}$ 与 $(\tau\beta)_{ij'}$ 的协方差为零, 随机误差项 ε_{ijk} 则是 $\text{NID}(0, \sigma^2)$. 因为交互作用效应在固定因子的各水平上的总和等于零, 这种混合模型也称为**约束模型**.

在这个模型中, $(\tau\beta)_{ij}$ 的方差定义为 $[(a-1)/a]\sigma_{\tau\beta}^2$, 而不是 $\sigma_{\tau\beta}^2$. 这是为了简化期望均方. 而假定 $(\tau\beta)_{\cdot j} = 0$ 对期望均方也有作用, 可以证明

$$\begin{aligned} &\text{E}(MS_A) = \sigma^2 + n\sigma_{\tau\beta}^2 + \frac{bn\sum\limits_{i=1}^{a}\tau_i^2}{a-1}, \qquad E(MS_{AB}) = \sigma^2 + n\sigma_{\tau\beta}^2 \\ &\text{E}(MS_B) = \sigma^2 + an\sigma_\beta^2, \qquad E(MS_E) = \sigma^2 \end{aligned} \tag{13.23}$$

因此, 检验固定因子效应的均值是否相等 (即 $H_0: \tau_i = 0$) 的恰当的统计量是

$$F_0 = \frac{MS_A}{MS_{AB}}$$

它服从 $F_{a-1,(a-1)(b-1)}$ 分布. 为检验 $H_0: \sigma_\beta^2 = 0$, 检验统计量是

$$F_0 = \frac{MS_B}{MS_E}$$

其分布为 $F_{b-1,ab(n-1)}$. 最后为检验 $H_0: \sigma_{\tau\beta}^2 = 0$, 用

$$F_0 = \frac{MS_{AB}}{MS_E}$$

其分布为 $F_{(a-1)(b-1),ab(n-1)}$.

在混合模型中, 可以估计固定效应为

$$\hat{\mu} = \bar{y}_{\cdots} \tag{13.24}$$

$$\hat{\tau}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots} \qquad i = 1, 2, \cdots, a$$

方差分量 $\sigma_\beta^2, \sigma_{\tau\beta}^2, \sigma^2$ 可以用方差分析法来估计. 在 (13.23) 式中去掉第一个方程, 留下有 3 个未知量的 3 个方程, 其解为

$$\hat{\sigma}_\beta^2 = \frac{MS_B - MS_E}{an}$$

$$\hat{\sigma}^2_{\tau\beta} = \frac{MS_{AB} - MS_E}{n} \tag{13.25}$$
$$\hat{\sigma}^2 = MS_E$$

这个一般的方法可以用来估计任一混合模型的方差分量. 在舍弃含有固定因子的均方后, 通常在留下的方程组中能够解出方差分量.

混合模型中, 试验者可能会对固定因子的某个处理均值的假设检验或置信区间构造感兴趣. 在使用这些方法时, 必须小心地使用处理均值的合适标准误. 固定效应处理均值的标准误是

$$\left[\frac{\text{用来检验固定效应的均方值}}{\text{每个处理均值的观测值个数}}\right]^{1/2} = \sqrt{\frac{MS_{AB}}{bn}}$$

注意该标准误就是将固定效应模型中使用的标准误中的 MS_E 换成了用于假设检验的均方.

例 13.3 测量系统的能力研究实验再探

重新考虑例 13.2 提到的量具 R&R 实验. 现假定只有 3 名操作员使用这种量具, 所以操作员是固定因子. 然而, 因为工件是随机选择的, 所以现在实验要采用混合模型.

混合模型的方差分析列在表 13.6 中. 计算由 Minitab 中的 Balanced ANOVA 过程完成. 我们规定在 Minitab 分析的模型是约束模型. Minitab 还生成了该模型的期望均方. 在 Minitab 的输出中, 值 Q[2] 标记了固定因子效应操作员的二次表达式. 即 Q[2]$=\sum_{j=1}^{b}\beta_j^2/(b-1)$. 结论与例 13.2 相似. 由 (13.25) 式, 方差分量可估计为

表 13.6 例 13.3 中混合模型的方差分析 (Minitab): 假定采用约束模型

```
Analysis of Variance (Balanced Designs)

Factor      Type Levels Values
part      random     20      1     2     3     4     5     6     7
                             8     9    10    11    12    13    14
                            15    16    17    18    19    20
operator fixed        3      1     2     3

Analysis of Variance for y

Source              DF          SS         MS        F       P
part                19    1185.425     62.391    62.92   0.000
operator             2       2.617      1.308     1.84   0.173
part*operator       38      27.050      0.712     0.72   0.861
Error               60      59.500      0.992
Total              119    1274.592

Source            Variance    Error   Expected Mean Square for Each Term
                  component    term   (using restricted model)
1 part             10.2332       4    (4) +  6(1)
2 operator                       3    (4) + 2(3) + 40Q[2]
3 part*operator    -0.1399       4    (4) + 2(3)
4 Error             0.9917            (4)
```

$$\hat{\sigma}^2_{\text{工件}} = \frac{MS_{\text{工件}} - MS_E}{an} = \frac{62.39 - 0.99}{(3)(2)} = 10.23$$

$$\hat{\sigma}^2_{\text{工件}\times\text{操作员}} = \frac{MS_{\text{工件}\times\text{操作员}} - MS_E}{n} = \frac{0.71 - 0.99}{2} = -0.14$$

$$\hat{\sigma}^2 = MS_E = 0.99$$

Minitab 输出也给出这些结果. 又得到了交互作用的方差分量的负的估计值. 一种合适的做法是, 正如我们在例 13.2 所做的, 拟合简化模型. 这时混合模型有两个因子, 得到与例 13.2 中相同的结果.

其他混合模型

人们曾提出几种不同的混合模型. 这些模型在关于随机分量的假定方面不同于上面研究的约束混合模型. 现简要讨论其中之一.

考虑模型

$$y_{ijk} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \varepsilon_{ijk}$$

其中 $\alpha_i(i = 1, 2, \cdots, a)$ 是满足 $\sum\limits_{i=1}^{a} \alpha_i = 0$ 的固定效应, γ_i、$(\alpha\gamma)_{ij}$ 以及 ε_{ijk} 是不相关的随机变量, 其均值都为零, 方差分别为 $\mathrm{V}(\gamma_j) = \sigma^2_\gamma$, $\mathrm{V}[(\alpha\gamma)_{ij}] = \sigma^2_{\alpha\gamma}$, $\mathrm{V}(\varepsilon_{ijk}) = \sigma^2$. 注意这里并没有用到原先对交互效应的约束条件, 这种混合模型一般被称为**无约束混合模型**.

可以证明, 这一模型的期望均方是 (参见本章的补充材料)

$$\begin{aligned} &\mathrm{E}(MS_A) = \sigma^2 + n\sigma^2_{\alpha\gamma} + \frac{bn\sum\limits_{i=1}^{a}\alpha_i^2}{a-1}, \qquad \mathrm{E}(MS_{AB}) = \sigma^2 + n\sigma^2_{\alpha\gamma} \\ &\mathrm{E}(MS_B) = \sigma^2 + n\sigma^2_{\alpha\gamma} + an\sigma^2_\gamma, \qquad \mathrm{E}(MS_E) = \sigma^2 \end{aligned} \tag{13.26}$$

将这些期望均方与 (13.23) 式中的期望均方进行比较, 可以看出仅有随机效应一项有明显的差别, 在随机效应的期望均方中出现方差分量 $\sigma^2_{\alpha\gamma}$. (实际上, 由于关于交互作用效应方差的定义不相同, 这两个模型还有其他的差别.) 因此, 我们用统计量 $F_0 = \frac{MS_B}{MS_{AB}}$ 检验随机效应的方差分量等于零 ($H_0 : \sigma^2_\gamma = 0$) 的这一假设类似于在约束模型中用 $F_0 = MS_B/MS_E$ 检验 $H_0 : \sigma^2_\beta = 0$.

两个模型的参数有密切的关系. 事实上, 可以证明

$$\tau_i = \alpha_i, \quad \beta_j = \gamma_j + (\overline{\alpha\gamma})_{\cdot j}, \quad (\tau\beta)_{ij} = (\alpha\gamma)_{ij} - (\overline{\alpha\gamma})_{\cdot j}$$

$$\sigma^2_\gamma = \sigma^2_\beta + \frac{1}{a}\sigma^2_{\alpha\gamma}, \quad \sigma^2_{\tau\beta} = \sigma^2_{\alpha\gamma}$$

可以用方差分析法来估计方差分量. 参照期望均方, 对 (13.25) 式仅需改变

$$\hat{\sigma}^2_\gamma = \frac{MS_B - MS_{\mathrm{AB}}}{an} \tag{13.27}$$

这两个模型都是 Scheffé (1956a, 1959) 所提出的混合模型的特殊情况. 该模型假定观测值可表示为

$$y_{ijk} = m_{ij} + \varepsilon_{ijk} \qquad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b \\ k = 1, 2, \cdots, n \end{cases}$$

其中 m_{ij} 与 ε_{ijk} 是独立随机变量. m_{ij} 的结构为

$$m_{ij} = \mu + \tau_i + b_j + c_{ij}, \quad \mathrm{E}(m_{ij}) = \mu + \tau_i, \quad \sum_{i=1}^{a}\tau_i = 0$$

$$c_{\cdot j}=0, \qquad j=1,2,\cdots,b$$

b_j 和 c_{ij} 的方差与协方差用 m_{ij} 的协方差来表示. 而且, 在混合模型的其他公式中, 随机效应参数和 b_j 与 c_{ij} 有关. 除了当 $H_0:\tau_i=0$ 为真时, 统计量 MS_A/MS_{AB} 不总是服从 F 分布这一点之外, Scheffé 模型的统计分析法和我们的约束模型相同.

鉴于混合模型的多样性, 一个逻辑上很自然的问题是, 人们该用哪一个模型呢? 我个人更看好约束模型, 而且它也是文献中最为常见的. 约束模型实际上比无约束模型更一般些. 因为在约束模型中, 随机因子水平相同的两个观测值的协方差可以是正的也可以是负的; 而在无约束模型中, 这种协方差只能是正的. 如果随机分量的相关结构不强, 则这些混合模型都适用, 它们之间仅有微小的差别. 另一方面, 混合模型的无约束形式更适用于设计为不平衡的场合, 因为它使用方便, 而且许多计算软件在显示期望均方时都假定为无约束模型. (SAS 就是如此, 而 Minitab 中的默认模型也是无约束模型, 尽管这很容易更改.) 以后, 当提到混合模型时, 我们假定是约束模型. 不过, 当数据有强的相关性时, 则必须用 Scheffé模型. 模型的选择应该主要取决于数据. Hocking (1973) 的论文清晰地总结了各种混合模型.

例 13.4 无约束模型

一些计算软件包只支持一种混合模型. 而 Mintab 同时支持约束与无约束模型, 尽管如上面提到的, 默认的是无约束模型. 表 13.7 列出了对例 13.3 中实验采用无约束模型的 Mintab 输出. 可以看到, 这些期望均方与 (13.26) 式中的期望均方相同. 结论和约束模型分析所得的相同, 而且方差分量估计也非常相近.

表 13.7 例 13.3 中采用无约束模型的实验分析

```
Analysis of Variance (Balanced Designs)

Factor     Type Levels Values
part      random     20     1    2    3    4    5    6    7
                            8    9   10   11   12   13   14
                           15   16   17   18   19   20
operator fixed        3     1    2    3

Analysis of Variance for y

Source            DF          SS         MS        F       P
part              19    1185.425     62.391    87.65   0.000
operator           2       2.617      1.308     1.84   0.173
part*operator     38      27.050      0.712     0.72   0.861
Error             60      59.500      0.992
Total            119    1274.592

Source            Variance   Error  Expected Mean Square for Each Term
                  component   term  (using unrestricted model)
1 part            10.2798      3    (4) + 2(3) +  6(1)
2 operator                     3    (4) + 2(3) +  Q[2]
3 part*operator   -0.1399      4    (4) + 2(3)
4 Error            0.9917           (4)
```

13.4 含随机效应的样本量的确定

附录中的抽检特性曲线可用于确定有随机因子实验的样本量. 考虑 13.1 节中的单因子随机效应模型. 该随机效应模型的第 II 类错误概率是

$$\beta = 1 - \mathrm{P}\{拒绝\ H_0|H_0不真\} = 1 - \mathrm{P}\{F_0 > F_{\alpha,a-1,N-a}|\sigma_\tau^2 > 0\} \tag{13.28}$$

这仍要用到检验统计量 $F_0 = MS_{处理}/MS_E$ 在备择假设下的分布. 可以证明, 当 H_1 为真 ($\sigma_\tau^2 > 0$) 时, F_0 的分布是自由度为 $a-1$ 和 $N-a$ 的中心 F 分布.

因为随机效应模型的第 II 类错误概率是基于常用的中心 F 分布, 我们可以用附录中的 F 分布表来求 (13.28) 式的值. 不过, 用抽检特性曲线来确定检验的灵敏度则更为简单. 附录的表 VI 给出了一系列有不同的分子自由度、分母自由度且 α 取 0.05 或 0.01 时的抽检特性曲线. 这类曲线画出关于参数 λ 的第 II 类错误的概率, 其中

$$\lambda = \sqrt{1 + \frac{n\sigma_\tau^2}{\sigma^2}} \tag{13.29}$$

注意 λ 含有两个未知参数 σ_τ^2 与 σ^2. 如果对全体处理中的有待检测的变异程度有所了解, 我们也许可以估计 σ_τ^2. σ^2 的估计可根据以往的经验与判断来选择. 有时, 通过探求比值 σ_τ^2/σ^2 将有助于我们确定所感兴趣的 σ_τ^2 的值.

例 13.5 假定有 5 个随机选出的处理, 每个处理有 6 个观测值, $\alpha = 0.05$, 若 σ_τ^2 等于 σ^2, 我们想要确定检验的势. 因为 $a = 5$, $n = 6$, 且 $\sigma_\tau^2 = \sigma^2$, 可算出

$$\lambda = \sqrt{1 + 6(1)} = 2.646$$

根据自由度为 $a - 1 = 4$, $N - a = 25$, 且 $\alpha = 0.05$ 的特性曲线, 可以得到

$$\beta \approx 0.20$$

所以检验的势近似为 0.80.

也可以通过观测标准差增加的百分比来确定样本量. 若各处理是齐次的, 则随机选取的观测的标准差为 σ. 但是, 若处理间有差异, 则随机选取的观测的标准差为

$$\sqrt{\sigma^2 + \sigma_\tau^2}$$

若 P 为给定的观测标准差的增加百分比, 超出它就要拒绝零假设,

$$\frac{\sqrt{\sigma^2 + \sigma_\tau^2}}{\sigma} = 1 + 0.01P$$

即

$$\frac{\sigma_\tau^2}{\sigma^2} = (1 + 0.01P)^2 - 1$$

因此, 由 (13.29) 式, 得到

$$\lambda = \sqrt{1 + \frac{n\sigma_\tau^2}{\sigma^2}} = \sqrt{1 + n[(1 + 0.01P)^2 - 1]} \tag{13.30}$$

对给定的 P, 附录的表 VI 中的特性曲线可用来确定所需的样本容量.

对于二因子随机效应模型和混合模型，我们仍可以用抽检特性曲线来确定样本的容量. 附录 VI 用于随机效应模型. 参数 λ、分子自由度、分母自由度列在表 13.8 的上半部分. 对于混合模型, 附录 V 与附录 VI 都要用到. 合适的 Φ^2 值与 λ 值列在表 13.8 的下半部分.

表 13.8 用于附录中表 V 与表 VI 中关于二因子随机效应模型和混合模型的抽检特性曲线的参数

随机效应模型			
因子	λ	分子自由度	分母自由度
A	$\sqrt{1+\frac{bn\sigma_\tau^2}{\sigma^2+n\sigma_{\tau\beta}^2}}$	$a-1$	$(a-1)(b-1)$
B	$\sqrt{1+\frac{an\sigma_\beta^2}{\sigma^2+n\sigma_{\tau\beta}^2}}$	$b-1$	$(a-1)(b-1)$
AB	$\sqrt{1+\frac{n\sigma_{\tau\beta}^2}{\sigma^2}}$	$(a-1)(b-1)$	$ab(n-1)$

混合模型				
因子	参数	分子自由度	分母自由度	附表
A(固定)	$\Phi^2=\frac{bn\sum_{i=1}^{a}\tau_i^2}{a[\sigma^2+n\sigma_{\tau\beta}^2]}$	$a-1$	$(a-1)(b-1)$	V
B(随机)	$\lambda=\sqrt{1+\frac{an\sigma_\beta^2}{\sigma^2}}$	$b-1$	$ab(n-1)$	VI
AB	$\lambda=\sqrt{1+\frac{n\sigma_{\tau\beta}^2}{\sigma^2}}$	$(a-1)(b-1)$	$ab(n-1)$	VI

13.5 期望均方的计算法则

实验设计问题的重要部分是进行方差分析. 这涉及要确定模型每一分量的平方和以及与每一平方和相关的自由度. 然后, 为了构造恰当的检验统计量, 就必须确定期望均方. 在复杂的设计情况下, 特别是涉及随机模型或混合模型的设计, 给这一过程建立一种规范的方法常常是有好处的.

我们将提出一组计算法则, 用来计算模型中各效应 (此处说的效应包括主效应与交互作用) 的平方和、自由度以及期望均方. 这些法则对任何平衡的析因设计、嵌套设计[①]或嵌套析因实验都适用. (但对部分的平衡安排, 如拉丁方与不完全区组设计, 并不适用.) 此外还有其他一些法则, 例如, 参见 Scheffé(1959), Bennett and Franklin(1954), Cornfield and Tukey(1956) 以及 Searle(1971a, 1971b). 通过考察期望均方, 人们可以提出检验关于任一模型参数的假设的恰当统计量. 检验统计量是期望均方的一个比值, 它的取法是使分子均方的期望值与分母均方的期望值仅仅在我们感兴趣的方差分量或固定因子上有所不同.

对任一模型, 通常有可能像我们在第 3 章中曾做过的那样去确定期望均方, 也就是, 通过直接应用期望算子. 这种常被叫做 "硬算" 的方法, 是很冗长乏味的. 下面所给出的求任一设计期

① 嵌套设计将在第 14 章介绍.

望均方的法则, 不需要硬算, 且在实践中, 用起来相对简单. 当应用于混合模型时, 我们用假定有 n 次重复的二因子固定效应析因模型来说明这些法则.

法则 1 模型中的误差项为 $\varepsilon_{ij\cdots m}$, 其中下标 m 是重复下标. 对二因子模型说来, 该法则意味着误差项为 ε_{ijk}. $\varepsilon_{ij\cdots m}$ 的方差分量是 σ^2.

法则 2 除了总均值 (μ) 和误差项 $\varepsilon_{ij\cdots m}$ 之外, 模型包含所有的主效应和任一实验者假定存在的交互作用. 如果 k 个因子所有可能的交互作用都存在, 则有 $\binom{k}{2}$ 个二因子交互作用, $\binom{k}{3}$ 个三因子交互作用,$\cdots$, 1 个 k 因子交互作用. 如果某一项中有一个因子出现在小括号内, 则在那一项中, 那个因子和其他的因子之间没有交互作用.

法则 3 对模型中除均值 (μ) 和误差项外的每一项, 将其下标划分为 3 类: (a) 可变动的 —— 在该项中出现的但不在小括号内出现的那些下标; (b) 固定的 —— 在该项中出现的且在小括号内出现的那些下标; (c) 不出现的 —— 在模型中出现但在该项中不出现的那些下标. 注意二因子固定效应模型中没有固定的下标, 不过我们将在以后碰到这种模型. 比如, 在二因子模型中, 对于 $(\tau\beta)_{ij}$, i 与 j 是可变动的, k 是不出现的.

法则 4 (自由度) 模型中任一项的自由度是与每一固定下标有关的水平数和与每一可变动下标有关的水平数减一的乘积. 例如, $(\tau\beta)_{ij}$ 的自由度是 $(a-1)(b-1)$. 误差的自由度是用 $N-1$ 减去所有其他自由度的和, 其中 N 是观测的总数.

法则 5 模型中的每一项都有一个与它有关的方差分量 (随机效应) 或固定因子 (固定效应). 如果某一交互作用至少包含一个随机效应时, 则整个交互作用看作是随机的. 方差分量用希腊字母作为下标来标记特定的随机效应. 这样, 在因子 A 固定而因子 B 随机的二因子混合模型中, B 的方差分量是 σ_β^2, 而 AB 的方差分量是 $\sigma_{\tau\beta}^2$. 固定效应通常用与那一因子有关的模型分量的平方和除以它的自由度来表示. 在我们的例子中, A 的固定效应是

$$\frac{\sum_{i=1}^{a} \tau_i^2}{a-1}$$

法则 6 (期望均方) 每一模型分量都有期望均方. 误差的期望均方为 $\mathrm{E}(MS_E)=\sigma^2$. 在**约束模型**中, 对模型的其他各项来讲, 期望均方或包含该项的方差分量或包含该项的固定效应分量, 加上问题中包含的效应的所有模型其他项的分量, 以及与其他固定效应没有交互作用的其他项的分量, 再加上 σ^2. 每个方差分量或固定效应的系数是在该分量取不同值时的观测个数.

用二因子固定效应模型的情形来说明, 考虑交互作用的均方期望 $\mathrm{E}(MS_{AB})$ 的计算. 均方期望只包含 AB 交互作用的固定效应 (因为没有其他模型项包含 AB) 和 σ^2. 而 AB 的固定效应要乘以 n, 因为在交互作用分量的每个不同值处有 n 个观测值 (在每个单元处有 n 个观测). 所以, AB 的均方期望为

$$\mathrm{E}(MS_{AB})=\sigma^2+\frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$$

作为二因子固定效应模型中的另一个说明, 主效应 A 的均方期望为

$$\mathrm{E}(MS_A)=\sigma^2+\frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$$

分子中的乘数是 bn, 因为 A 的每个水平上有 bn 个观测. 交互作用项 AB 没有包含在期望均方中, 因为虽然它的确包含与 A 有关的效应, 但是 B 是固定效应.

为说明法则 6 如何用于随机效应模型, 我们考虑二因子随机模型. 交互作用 AB 的均方期望为

$$E(MS_{AB}) = \sigma^2 + n\sigma_{\tau\beta}^2$$

而主效应 A 的均方期望为

$$E(MS_A) = \sigma^2 + n\sigma_{\tau\beta}^2 + bn\sigma_\tau^2$$

注意这包含了交互作用 AB 的方差分量, 因为 A 包含在 AB 中而 B 是随机效应.

再考虑约束的二因子混合模型. 交互作用 AB 的均方期望为

$$E(MS_{AB}) = \sigma^2 + n\sigma_{\tau\beta}^2$$

而固定因子 A 的主效应的均方期望为

$$E(MS_A) = \sigma^2 + n\sigma_{\tau\beta}^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$$

它包含了交互作用的方差分量, 因为 A 包含在 AB 中而 B 是随机效应. 对于因子 B 的主效应, 期望均方为

$$E(MS_B) = \sigma^2 + an\sigma_\beta^2$$

这里交互作用的方差分量没有出现, 因为虽然 B 包含在 AB 中但 A 是固定效应. 请注意这些期望均方与前面 (12.23) 式的二因子混合模型的那些结果是一致的.

法则 6 可以很方便地修改, 以用于无约束混合模型的期望均方的计算. 仅仅包括所考虑项的效应, 加上包含该效应的所有项, (只要存在至少一个随机因子). 为了说明, 考虑无约束二因子混合模型的情况. 二因子的交互作用项的均方期望为

$$E(MS_{AB}) = \sigma^2 + n\sigma_{\tau\beta}^2$$

(请回忆约束模型与无约束模型中模型分量记号的差别.) 对于固定因子 A 的主效应, 期望均方为

$$E(MS_A) = \sigma^2 + n\sigma_{\tau\beta}^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$$

而对于随机因子 B 的主效应, 期望均方则为

$$E(MS_B) = \sigma^2 + n\sigma_{\tau\beta}^2 + an\sigma_\beta^2$$

注意这些是原先关于无约束混合模型的 (13.26) 式给出的期望均方.

例 13.6 考虑三因子析因实验, 因子 A 有 a 个水平, 因子 B 有 b 个水平, 因子 C 有 c 个水平, n 次重复. 假定所有因子都是固定效应, 5.4 节给出了这一设计的分析. 现在假定所有因子都是随机的, 要确定期望均方, 合适的统计模型是

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

用前面所说的法则, 期望均方的演算列在表 13.9 中.

由表 13.9 的期望均方, 我们注意到, 如果 A, B, C 都是随机因子, 则对主效应不存在精确的检验法. 也就是说, 如果想要检验假设 $\sigma_\tau^2=0$, 则我们不能构造两个期望均方的比值使得分子仅比分母多一项 $bcn\sigma_\tau^2$. B 与 C 的主效应也有同样的现象. 对二因子和三因子的交互作用的精确检验法是存在的. 但是, 对实验者来说, 关于主效应的检验法可能更加重要, 因此, 应该怎样来检验主效应呢? 这一问题在 13.6 节中考虑.

表 13.9 三因子随机效应模型的期望均方

模型项	因 子	期望均方
τ_i	A, 主效应	$\sigma^2+cn\sigma_{\tau\beta}^2+bn\sigma_{\tau\gamma}^2+n\sigma_{\tau\beta\gamma}^2+bcn\sigma_\tau^2$
β_j	B, 主效应	$\sigma^2+cn\sigma_{\tau\beta}^2+an\sigma_{\beta\gamma}^2+n\sigma_{\tau\beta\gamma}^2+acn\sigma_\beta^2$
γ_k	C, 主效应	$\sigma^2+bn\sigma_{\tau\gamma}^2+an\sigma_{\beta\gamma}^2+n\sigma_{\tau\beta\gamma}^2+abn\sigma_\gamma^2$
$(\tau\beta)_{ij}$	AB, 二因子交互作用	$\sigma^2+n\sigma_{\tau\beta\gamma}^2+cn\sigma_{\tau\beta}^2$
$(\tau\gamma)_{ik}$	AC, 二因子交互作用	$\sigma^2+n\sigma_{\tau\beta\gamma}^2+bn\sigma_{\tau\gamma}^2$
$(\beta\gamma)_{jk}$	BC, 二因子交互作用	$\sigma^2+n\sigma_{\tau\beta\gamma}^2+an\sigma_{\beta\gamma}^2$
$(\tau\beta\gamma)_{ijk}$	ABC, 三因子交互作用	$\sigma^2+n\sigma_{\tau\beta\gamma}^2$
ϵ_{ijkl}	误差	σ^2

13.6 近似 F 检验

在有 3 个或更多个因子的随机模型或混合模型的析因实验中, 或某些更为复杂的设计中, 模型的某些效应常常没有精确的检验统计量. 应对这种困境的一种可能的解法是假定某些交互作用可以忽略. 比如, 当有理由假定例 13.6 中所有的二因子交互作用都可以忽略时, 则可令 $\sigma_{\tau\beta}^2=\sigma_{\tau\gamma}^2=\sigma_{\beta\gamma}^2=0$, 从而可进行主效应的检验.

虽然这样做看起来很有吸引力, 但必须指出, 为了假定有一个或多个交互作用可以忽略, 须要对生产过程的本质有所了解或者说要有一些坚实的先验知识才行. 一般说来, 这种假定是不易做到的, 也不应该轻易使用. 在没有足够的证据证明那样做是恰当的时候, 不应当删除模型的某些交互作用. 有些实验者提倡的方法是首先检验交互作用, 然后令那些不显著的交互作用为零, 再在同一实验中检验其他效应时假定这些交互作用为零. 尽管有时在实践中会这样做, 但这一方法是有危险的, 因为关于交互作用的任何决定, 都使人们容易遭受第 I 类和第 II 类错误.

这一想法的一种变形, 是在方差分析中将某些均方**合并**起来, 以便得到有更大自由度的误差估计量. 例如, 在例 13.6 中, 假定检验统计量 $F_0=MS_{ABC}/MS_E$ 不显著. 于是, $H_0:\sigma_{\tau\beta\gamma}^2=0$ 不能被否定, 因此, MS_{ABC} 与 MS_E都可估计误差方差 σ^2. 实验者可按照下式合并 MS_{ABC} 与 MS_E:

$$MS_{E'}=\frac{abc(n-1)MS_E+(a-1)(b-1)(c-1)MS_{ABC}}{abc(n-1)+(a-1)(b-1)(c-1)}$$

使得 $\mathrm{E}(MS_{E'})=\sigma^2$. $MS_{E'}$ 有 $abc(n-1)+(a-1)(b-1)(c-1)$ 个自由度, 而原先的 MS_E 有 $abc(n-1)$ 个自由度.

合并的危险在于, 可能产生第 II 类错误并将实际上有显著性的因子的均方与误差合并起来, 从而导致新的残差均方 ($MS_{E'}$) 太大. 这会使其他显著性效应更难于检测. 另一方面, 如果原来

的误差均方的自由度很小 (例如, 小于 6), 由于合并可能会潜在地增加进一步检验的精确度, 实验者也许会获益更多. 一种合理的实用方法如下. 如果原来的误差均方的自由度等于 6 或更多, 则不合并, 如果原来的误差均方的自由度小于 6, 则仅当被合并的均方的 F 统计量在大的 α 值 (比方说 $\alpha = 0.25$) 处不显著时才进行合并.

如果不能假定某些交互作用可忽略, 而我们仍须作出关于那些不存在精确检验法的效应的统计推断时, 则可以使用 Satterthwaite (1946) 所提供的方法. Satterthwaite 法使用**均方的线性组合**, 例如,

$$MS' = MS_r + \cdots + MS_s \tag{13.31}$$

$$MS'' = MS_u + \cdots + MS_v \tag{13.32}$$

其中 (13.31) 式与 (13.32) 式的均方选择是要使得 $E(MS') - E(MS'')$ 等于零假设中所考虑的效应 (模型参数或方差分量) 的倍数. 此时检验统计量是

$$F = \frac{MS'}{MS''} \tag{13.33}$$

其分布近似为 $F_{p,q}$, 其中

$$p = \frac{(MS_r + \cdots + MS_s)^2}{MS_r^2/f_r + \cdots + MS_s^2/f_s} \tag{13.34}$$

$$q = \frac{(MS_u + \cdots + MS_v)^2}{MS_u^2/f_u + \cdots + MS_v^2/f_v} \tag{13.35}$$

在 p 与 q 中, f_i 是均方 MS_i 的自由度. 我们不能确保 p 与 q 都是整数, 所以需要在 F 分布表中进行插值. 例如, 在三因子随机效应模型 (表 13.9) 中, 易见用于检验 $H_0: \sigma_\tau^2 = 0$ 的合适的检验统计量是 $F = MS'/MS''$, 其中

$$MS' = MS_A + MS_{ABC}, \qquad MS'' = MS_{AB} + MS_{AC}$$

F 的自由度可以由 (13.34) 和 (13.35) 式算得.

该检验的理论依据是检验统计量 [(13.33) 式] 的分子与分母都近似为卡方变量的乘积, 而且因为没有均方同时出现在分子与分母中, 所以分子与分母相互独立, 所以 (13.33) 式中的 F 近似服从 $F_{p,q}$ 分布. Satterthwaite 指出, 当 MS' 与 MS'' 中某些均方取负值时, 应用上述方法要特别小心. Gaylor and Hopper(1969) 提出, 若 $MS' = MS_1 - MS_2$, 则当

$$\frac{MS_1}{MS_2} > F_{0.025,f_2,f_1} \times F_{0.50,f_2,f_2}$$

且 $f_1 \leqslant 100$ 和 $f_2 \geqslant f_1/2$ 时, Satterthwaite 的近似检验仍然适用.

例 13.7 研究涡轮机膨胀阀的压力差实验. 设计工程师考虑的影响压力差的重要变量有: 入口端的气体温度 (A)、测量员 (B)、测量员所用的测压表 (C). 这 3 个因子被安排在一个析因设计中, 其中气温固定, 测量员与测压表是随机的. 表 13.10 中给出有两次重复的规范数据. 该设计的线性模型是:

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

其中 τ_i 是气体温度 (A) 的效应, β_j 是测量员 (B) 的效应, γ_k 是测压表 (C) 的效应.

方差分析见表 13.11. 期望均方的一列加进该表中, 该列中的各项都是用 13.5 节所讨论的准则推导出来的. 由均方期望这一列可看到, 除 A 的主效应外, 对其他所有效应都存在精确的检验法. 检验结果显示在表 13.11 中.

要检验气体温度效应, 或 $H_0: \tau_i = 0$, 使用统计量

$$F = \frac{MS'}{MS''}$$

表 13.10　涡轮机实验的压力差规范数据

测压表	温度 (A)											
	60°F				75°F				90°F			
	测量员 (B)				测量员 (B)				测量员 (B)			
(C)	1	2	3	4	1	2	3	4	1	2	3	4
1	−2	0	−1	4	14	6	1	−7	−8	−2	−1	−2
	−3	−9	−8	4	14	0	2	6	−8	20	−2	1
2	−6	−5	−8	−3	22	8	6	−5	−8	1	−9	−8
	4	−1	−2	−7	24	6	2	2	3	−7	−8	3
3	−1	−4	0	−2	20	2	3	−5	−2	−1	−4	1
	−2	−8	−7	4	16	0	0	−1	−1	−2	−7	3

表 13.11　压力差数据的方差分析

方差来源	平方和	自由度	期望均方	均方	F_n	P 值
温度, A	1 023.36	2	$\sigma^2 + bn\sigma^2_{\tau\gamma} + cn\sigma^2_{\tau\beta} + n\sigma^2_{\tau\beta\gamma} + \frac{bcn\sum\tau_i^2}{a-1}$	511.68	2.22	0.17
测量员, B	423.82	3	$\sigma^2 + an\sigma^2_{\beta\gamma} + acn\sigma^2_{\beta}$	141.27	4.05	0.07
测压表, C	7.19	2	$\sigma^2 + an\sigma^2_{\beta\gamma} + abn\sigma^2_{\gamma}$	3.60	0.10	0.90
AB	1 211.97	6	$\sigma^2 + n\sigma^2_{\tau\beta\gamma} + cn\sigma^2_{\tau\beta}$	202.00	14.59	<0.01
AC	137.89	4	$\sigma^2 + n\sigma^2_{\tau\beta\gamma} + bn\sigma^2_{\tau\gamma}$	34.47	2.49	0.10
BC	209.47	6	$\sigma^2 + an\sigma^2_{\beta\gamma}$	34.91	1.63	0.17
ABC	166.11	12	$\sigma^2 + n\sigma^2_{\tau\beta\gamma}$	13.84	0.65	0.79
误差	770.50	36	σ^2	21.40		
合计	3 950.32	71				

其中

$$MS' = MS_A + MS_{ABC}, \qquad MS'' = MS_{AB} + MS_{AC}$$

因为

$$E(MS') - E(MS'') = \frac{bcn\sum\tau_i^2}{a-1}$$

为确定检验 $H_0: \tau_i = 0$ 的统计量, 我们算得

$$MS' = MS_A + MS_{ABC} = 511.68 + 13.84 = 525.52$$
$$MS'' = MS_{AB} + MS_{AC} = 202.00 + 34.47 = 236.47$$
$$F = \frac{MS'}{MS''} = \frac{525.52}{236.47} = 2.22$$

根据 (13.34) 和 (13.35) 式, 此统计量的自由度如下:

$$p = \frac{(MS_A + MS_{ABC})^2}{MS_A^2/2 + MS_{ABC}^2/12} = \frac{(525.52)^2}{(511.68)^2/2 + (13.84)^2/12} = 2.11 \approx 2$$

$$q=\frac{(MS_{AB}+MS_{AC})^2}{MS_{AB}^2/6+MS_{AC}^2/4}=\frac{(236.47)^2}{(202.00)^2/6+(34.47)^2/4}=7.88\approx 8$$

将 $F=2.22$ 与 $F_{0.05,2,8}=4.46$ 相比较，我们不能拒绝 H_0. 其 P 值约等于 $P=0.17$.

AB(即气温与测量员) 的交互作用是大的，而有迹象表明 AC(即气温与测压表) 存在交互作用. AB 交互作用与 AC 交互作用的图形分析显示在图 13.2 中，它表明当使用测量员 1 和测量表 3 时，气温的效应比较大. 这样一来，气温与测量员的主效应有可能被大的 AB 交互作用所掩盖.

表 13.12 是 Minitab Balanced ANOVA 关于例 13.7 中的实验的输出结果. 我们规定的是**约束模型**, Q[1] 代表了气压的固定效应. 可以看到，方差分析表中的各项与表 13.11 中的基本相同，除了气温的 F 检验统计量的值 (因子 A) 以外. Minitab 注明该检验不是精确检验 (由期望均方可见). Minitab 用的合成检验就是 Satterthwaite 方法，但它的检验统计量与我们的有点不同. 从 Minitab 的输出结果可以看到，检验因子 A 的误差均方是

$$(4)+(5)-(7)=MS_{AB}+MS_{AC}-MS_{ABC}$$

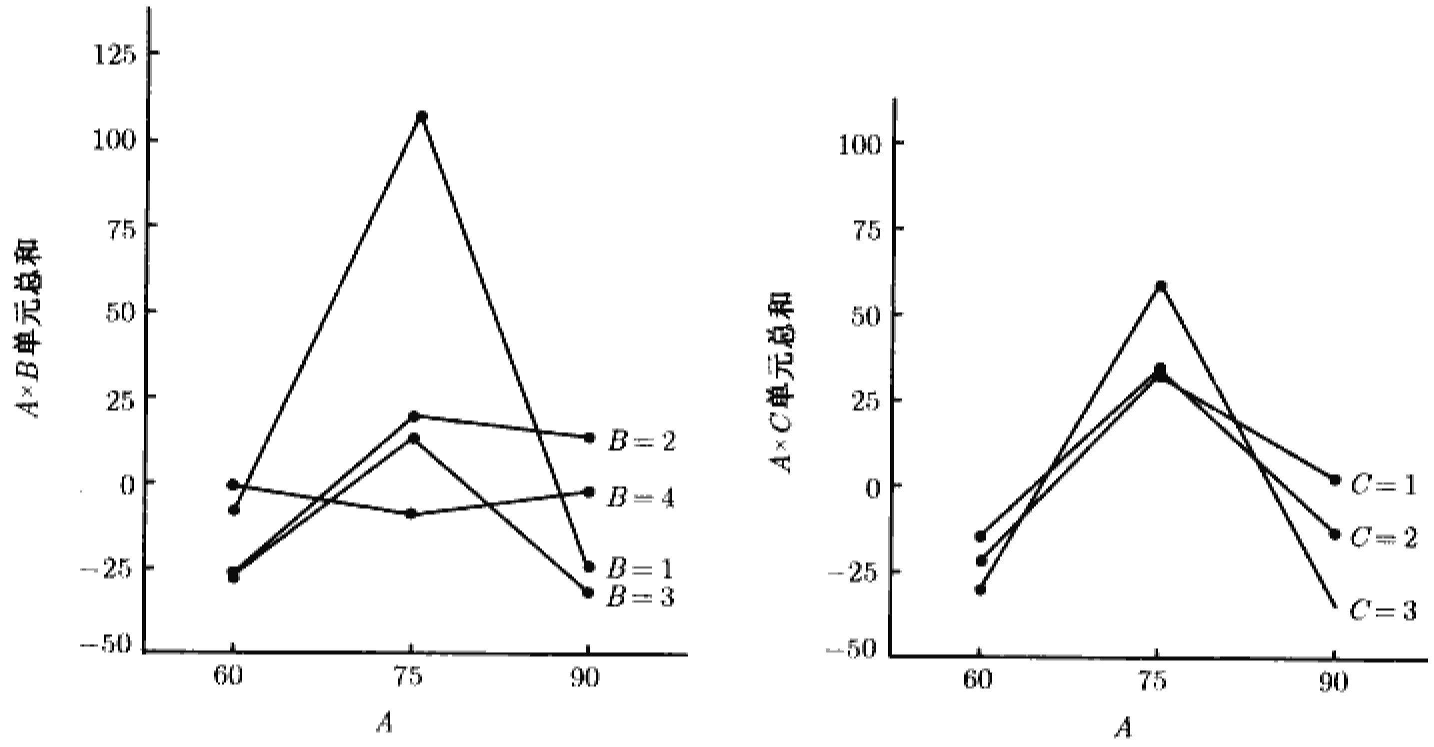


图 13.2 压力差实验中的交互作用

其期望值为：

$$\begin{aligned}E[(4)+(5)-(7)]&=\sigma^2+n\sigma^2_{\tau\beta\gamma}+cn\sigma^2_{\tau\beta}+\sigma^2+n\sigma^2_{\tau\beta\gamma}+bn\sigma^2_{\tau\gamma}-(\sigma^2+n\sigma^2_{\tau\beta\gamma})\\&=\sigma^2+n\sigma^2_{\tau\beta\gamma}+cn\sigma^2_{\tau\beta}+bn\sigma^2_{\tau\gamma}\end{aligned}$$

它是一个合适的用来检验 A 的均值效应的近似误差均方. 这很好地说明了 Satterthwaite 方法中可以用不止一种的方法来构造合成均方. 不过，一般我们更倾向于选择均方的线性组合，而不是 Minitab 所用的那种线性组合，因为它不能保证该线性组合的均方非负.

表 13.13 给出了例 13.7 的无约束模型的分析. 与约束模型的主要差别在于，对全部 3 个均值效应的均方的期望值是不存在精确检验的. 无约束模型中两个随机均值效应可以用它们的交互效应进行检验，但这时 B 的均方期望涉及 $\sigma^2_{\tau\beta\gamma}$ 和 $\sigma^2_{\tau\beta}$, C 的均方期望涉及 $\sigma^2_{\tau\beta\gamma}$ 和 $\sigma^2_{\tau\gamma}$. Minitab 又一次构造合成的均方，并用 Satterthwaite 方法来检验这些效应. 除了测量员效应的

方差外，无约束模型的结论与约束模型的大体上没有根本的差异. 无约束模型得到了一个负的 σ_β^2 的估计. 因为测量表因子在两个分析中都不显著，接下来进行某种模型简化是可能的.

表 13.12　例 13.7 的 Minitab Balanced ANOVA 输出 (约束模型)

Analysis of Variance (Balanced Designs)

Factor	Type	Levels	Values			
GasT	fixed	3	60	75	90	
Operator	random	4	1	2	3	4
Gauge	random	3	1	2	3	

Analysis of Variance for Drop

Source	DF	SS	MS	F	P	
GasT	2	1023.36	511.68	2.30	0.171	x
Operator	3	423.82	141.27	4.05	0.069	
Gauge	2	7.19	3.60	0.10	0.904	
Gast*Operator	6	1211.97	202.00	14.59	0.000	
Gast*Gauge	4	137.89	34.47	2.49	0.099	
Operator*Gauge	6	209.47	34.91	1.63	0.167	
Gast*Operator*Gauge	12	166.11	13.84	0.65	0.788	
Error	36	770.50	21.40			
Total	71	3950.32				

x Not an exact F-test.

Source	Variance component	Error term	Expected Mean Square for Each Term (using restricted model)
1 GasT		*	(8) + 2(7) + 8(5) + 6(4) + 24Q[1]
2 Operator	5.909	6	(8) + 6(6) + 18(2)
3 Gauge	-1.305	6	(8) + 6(6) + 24(3)
4 GasT*Operator	31.559	7	(8) + 2(7) + 6(4)
5 GasT*Gauge	2.579	7	(8) + 2(7) + 8(5)
6 Operator*Gauge	2.252	8	(8) + 6(6)
7 Gast*Operator*Gauge	-3.780	8	(8) + 2(7)
8 Error	21.403		(8)

* Synthesized Test.

Error Terms for Synthesized Tests

Source	Error DF	Error MS	Synthesis of Error MS
1 GasT	6.97	222.63	(4) + (5) - (7)

表 13.13 例 13.7 的 Minitab Balanced ANOVA 输出 (无约束模型)

```
Analysis of Variance (Balanced Designs)

Factor          Type   Levels   Values
GasT           fixed        3       60       75       90
Operator      random        4        1        2        3       4
Gauge         random        3        1        2        3

Analysis of Variance for Drop

Source                 DF        SS        MS        F        P
GasT                    2   1023.36    511.68     2.30    0.171  x
Operator                3    423.82    141.27     0.63    0.616  x
Gauge                   2      7.19      3.60     0.06    0.938  x
Gast*Operator           6   1211.97    202.00    14.59    0.000
Gast*Gauge              4    137.89     34.47     2.49    0.099
Operator*Gauge          6    209.47     34.91     2.52    0.081
Gast*Operator*Gauge    12    166.11     13.84     0.65    0.788
Error                  36    770.50     21.40 Total 71  3950.32

x Not an exact F-test.

Source                 Variance  Error  Expected Mean Square for Each Term
                       component term   (using unrestricted model)
1 GasT                             *    (8) + 2(7) + 8(5) + 6(4) + Q[1]
2 Operator               -4.544    *    (8) + 2(7) + 6(6) + 6(4) + 18(2)
3 Gauge                  -2.164    *    (8) + 2(7) + 6(6) + 8(5) + 24(3)
4 GasT*Operator          31.359    7    (8) + 2(7) + 6(4)
5 GasT*Gauge              2.579    7    (8) + 2(7) + 8(5)
6 Operator*Gauge          3.512    7    (8) + 2(7) + 6(6)
7 Gast*Operator*Gauge    -3.780    8    (8) + 2(7)
8 Error                  21.403         (8)

* Synthesized Test.

Error Terms for Synthesized Tests

Source             Error DF    Error MS     Synthesis of Error MS
1 GasT                 6.97      222.63     (4) + (5) - (7)
2 Operator             7.09      223.06     (4) + (6) - (7)
3 Gauge                5.98       55.54     (5) + (6) - (7)
```

13.7 关于方差分量估计的一些其他论题

正如我们前面看到的, 随机模型或混合模型中方差分量的估计, 对实验者来讲, 常常是非常重要的目标. 本节将给出一些在估计方差分量时更深入的结果和更有用的技术. 主要考虑计算方差分量的置信区间的方法, 以及说明如何计算方差分量的最大似然估计. 当方差分析方法产生负的估计时, 最大似然估计方法可能是非常有用的替换方法.

13.7.1 方差分量的近似置信区间

13.1 节引入了随机效应模型, 当时给出了简单设计中 σ^2 以及方差分量的其他函数的精确 $100(1-\alpha)\%$ 置信区间. 方差分析中, 对任意一个能表示为某个均方期望的方差分量的函数, 总能找到它的精确置信区间. 例如, 考虑误差均方. 因为 $\mathrm{E}(MS_E)=\sigma^2$, 我们可以求出 σ^2 的精确置信区间, 这是因为

$$f_E MS_E/\sigma^2 = f_E\hat{\sigma}^2/\sigma^2$$

服从自由度为 f_E 的卡方分布. 精确的 $100(1-\alpha)\%$ 置信区间是

$$\frac{f_E MS_E}{\chi^2_{\alpha/2,f_E}} \leqslant \sigma^2 \leqslant \frac{f_E MS_E}{\chi^2_{1-\alpha/2,f_E}} \tag{13.36}$$

不幸的是, 在很多涉及多个设计因子的复杂实验中, 一般不能找到那些方差分量的精确置信区间, 这是因为这种方差不是方差分析中单个均方的精确期望. 不过, 在 13.6 节中介绍的 Satterthwaite 的近似伪 F 检验中的概念, 可以用于对那些没有精确置信区间的方差分量构造近似置信区间.

Satterthwaite 方法中用到均方的两个线性组合:

$$MS' = MS_r + \cdots + MS_s$$

和

$$MS'' = MS_u + \cdots + MS_v$$

检验统计量

$$F = \frac{MS'}{MS''}$$

近似于 F 分布. 采用 (13.34) 和 (13.35) 式定义的 MS' 与 MS'' 的近似自由度, 可以将这个伪 F 统计量用于对感兴趣的参数或方差分量的显著性的近似检验.

要检验方差分量 σ_0^2 的显著性, 选择两个线性组合 MS' 与 MS'', 使它们期望值的差等于该方差分量的某个数乘, 即

$$\mathrm{E}(MS') - \mathrm{E}(MS'') = k\sigma_0^2$$

或

$$\sigma_0^2 = \frac{\mathrm{E}(MS') - \mathrm{E}(MS'')}{k} \tag{13.37}$$

(13.37) 式给出了 σ_0^2 的一个点估计:

$$\hat{\sigma}_0^2 = \frac{MS' - MS''}{k} = \frac{1}{k}MS_r + \cdots + \frac{1}{k}MS_s - \frac{1}{k}MS_u - \cdots - \frac{1}{k}MS_v \tag{13.38}$$

(13.38) 式中的各均方 (MS_i) 相互独立, 且 $f_i MS_i/\sigma_i^2 = SS_i/\sigma_i^2$ 有自由度为 f_i 的卡方分布. 方差分量 σ_0^2 的估计量是均方的一个线性组合, 而 $r\hat{\sigma}_0^2/\sigma_0^2$ 有自由度为 r 的近似卡方分布, 其中

$$r = \frac{(\hat{\sigma}_0^2)^2}{\sum_{i=1}^{m} \frac{1}{k^2}\frac{MS_i^2}{f_i}} = \frac{(MS_r + \cdots + MS_s - MS_u - \cdots - MS_v)^2}{\frac{MS_r^2}{f_r} + \cdots + \frac{MS_s^2}{f_s} + \frac{MS_u^2}{f_u} + \cdots + \frac{MS_v^2}{f_v}} \tag{13.39}$$

该结果仅在 $\sigma_0^2 > 0$ 时才成立. 因为 r 通常不是整数, 一般要用到卡方表的插值. Graybill(1961) 导出了关于 r 的一般结果.

因为 $r\hat{\sigma}_0^2/\sigma_0^2$ 具有自由度为 r 的近似卡方分布, 所以

$$\mathrm{P}\left\{\chi^2_{1-\alpha/2,r} \leqslant \frac{r\hat{\sigma}_0^2}{\sigma_0^2} \leqslant \chi^2_{\alpha/2,r}\right\} = 1-\alpha$$

即

$$\mathrm{P}\left\{\frac{r\hat{\sigma}_0^2}{\chi^2_{\alpha/2,r}} \leqslant \sigma_0^2 \leqslant \frac{r\hat{\sigma}_0^2}{\chi^2_{1-\alpha/2,r}}\right\} = 1-\alpha$$

所以 σ_0^2 的近似 $100(1-\alpha)\%$ 置信区间为

$$\frac{r\hat{\sigma}_0^2}{\chi^2_{\alpha/2,r}} \leqslant \sigma_0^2 \leqslant \frac{r\hat{\sigma}_0^2}{\chi^2_{1-\alpha/2,r}} \tag{13.40}$$

例 13.8 为说明该方法, 我们考虑例 13.7 中的实验, 这里用一个三因子混合模型研究涡轮机膨胀阀的压力差. 模型是

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

其中 τ_i 是固定效应, 其余效应都是随机效应. 我们要求出 $\sigma^2_{\tau\beta}$ 的近似置信区间. 根据表 13.11 中的均方期望可以看到, 二因子交互效应 AB 与三因子交互效应 ABC 上的均方期望的差是所感兴趣的方差分量 $\sigma^2_{\tau\beta}$ 的倍数:

$$\mathrm{E}(MS_{AB}) - \mathrm{E}(MS_{ABC}) = \sigma^2 + n\sigma^2_{\tau\beta\gamma} + cn\sigma^2_{\tau\beta} - (\sigma^2 + n\sigma^2_{\tau\beta\gamma}) = cn\sigma^2_{\tau\beta}$$

这样, $\sigma^2_{\tau\beta}$ 的点估计为

$$\hat{\sigma}^2_{\tau\beta} = \frac{MS_{AB} - MS_{ABC}}{cn} = \frac{202.00 - 13.84}{(3)(2)} = 31.36$$

且

$$r = \frac{(MS_{AB} - MS_{ABC})^2}{\frac{MS_{AB}^2}{(a-1)(b-1)} + \frac{MS_{ABC}^2}{(a-1)(b-1)(c-1)}} = \frac{(202.00 - 13.84)^2}{\frac{(202.00)^2}{(2)(3)} + \frac{(13.84)^2}{(2)(3)(2)}} = 5.19$$

由 (13.40) 式, 可得 $\sigma^2_{\tau\beta}$ 的近似 95%置信区间如下:

$$\frac{r\hat{\sigma}^2_{\tau\beta}}{\chi^2_{0.025,r}} \leqslant \sigma^2_{\tau\beta} \leqslant \frac{r\hat{\sigma}^2_{\tau\beta}}{\chi^2_{0.975,r}}$$

$$\frac{(5.19)(31.36)}{13.14} \leqslant \sigma^2_{\tau\beta} \leqslant \frac{(5.19)(31.36)}{0.90}$$

$$12.39 \leqslant \sigma^2_{\tau\beta} \leqslant 180.84$$

13.7.2　修正的大样本方法

上面的 Satterthwaite 方法是相对简单的构造方差分量近似置信区间的方法，所处理的方差分量能表示为均方的线性组合，即

$$\hat{\sigma}_0^2=\sum_{i=1}^{Q}c_iMS_i \tag{13.41}$$

当每个均方 MS_i 的自由度都比较大，并且 (13.41) 式中的所有 c_i 都为正数时，Satterthwaite 方法非常适用. 但是，经常会有一些 c_i 是负的. Graybill and Wang (1980) 提出一个称为**修正的大样本方法**，它是 Satterthwaite 方法的一个非常有用的补充. 若 (13.41) 式中的所有 c_i 都为正数，则 σ_0^2 的近似 $100(1-\alpha)\%$ 的大样本置信区间为

$$\hat{\sigma}_0^2-\sqrt{\sum_{i=1}^{Q}G_i^2c_i^2MS_i^2}\leqslant\sigma_0^2\leqslant\hat{\sigma}_0^2+\sqrt{\sum_{i=1}^{Q}H_i^2c_i^2MS_i^2} \tag{13.42}$$

其中

$$G_i=1-\frac{1}{F_{\alpha,f_i,\infty}},\qquad H_i=\frac{1}{F_{1-\alpha,f_i,\infty}}-1$$

注意，分母自由度是无穷的 F 随机变量等价于一个卡方分布的随机变量除以自身的自由度.

现在考虑比 (13.41) 式更一般的情况，其中的常数 c_i 没有都为正数的约束. 这可以表示为

$$\hat{\sigma}_0^2=\sum_{i=1}^{P}c_iMS_i-\sum_{j=P+1}^{Q}c_jMS_j,\qquad c_i,c_j\geqslant 0 \tag{13.43}$$

Ting 等人 (1990) 给出了 σ_0^2 的近似 $100(1-\alpha)\%$ 的置信下限为

$$L=\hat{\sigma}_0^2-\sqrt{V_L} \tag{13.44}$$

其中

$$\begin{aligned}
V_L=&\sum_{i=1}^{P}G_i^2c_i^2MS_i^2+\sum_{j=P+1}^{Q}H_j^2c_j^2MS_j^2+\sum_{i=1}^{P}\sum_{j=P+1}^{Q}G_{ij}^2c_ic_jMS_iMS_j\\
&+\sum_{i=1}^{P-1}\sum_{t>1}^{P}G_{it}^*c_ic_tMS_iMS_t\\
G_i=&1-\frac{1}{F_{\alpha,f_i,\infty}}\\
H_j=&\frac{1}{F_{1-\alpha,f_j,\infty}}-1\\
G_{ij}=&\frac{(F_{\alpha,f_i,f_j}-1)^2-G_i^2F_{\alpha,f_i,f_j}^2-H_j^2}{F_{\alpha,f_i,f_j}}\\
G_{it}^*=&\begin{cases}\left[\left(\dfrac{1}{F_{\alpha,f_i+f_t,\infty}}\right)^2\dfrac{(f_i+f_t)^2}{f_if_t}-\dfrac{G_i^2f_i}{f_t}-\dfrac{G_i^2f_i}{f_t}\right](P-1), & \text{当 } P>1\\ 0, & \text{当 } P=1\end{cases}
\end{aligned}$$

这些结果还可以推广到方差分量比值的近似置信区间. 关于这些方法的完整说明，参见 Burdick and Graybill(1992) 这一优秀专著. 也可以参见本章的补充材料.

例 13.9 为说明修正的大样本方法, 重新考虑例 13.7 中的三因子混合模型. 我们要求出 $\sigma^2_{\tau\beta}$ 的近似 95%置信区间. 回顾 $\sigma^2_{\tau\beta}$ 的点估计为

$$\hat{\sigma}^2_{\tau\beta} = \frac{MS_{AB} - MS_{ABC}}{cn} = \frac{202.00 - 13.84}{(3)(2)} = 31.36$$

因此, 用 (13.43) 式的记号, $c_1 = c_2 = \frac{1}{6}$, 且

$$G_1 = 1 - \frac{1}{F_{0.05,6,\infty}} = 1 - \frac{1}{2.1} = 0.524$$

$$H_2 = \frac{1}{F_{0.95,12,\infty}} - 1 = \frac{1}{0.435} - 1 = 1.30$$

$$G_{12} = \frac{(F_{0.05,6,12} - 1)^2 - (G_1)^2 F^2_{0.05,6,12} - (H_2)^2}{F_{0.05,6,12}}$$

$$= \frac{(3.00 - 1)^2 - (0.524)^2(3.00)^2 - (1.3)^2}{3.00} = -0.054$$

$$G^*_{1t} = 0$$

由 (13.44) 式,

$$\begin{aligned} \mathrm{V}_L &= G_1^2 c_1^2 MS_{AB}^2 + H_2^2 c_2^2 MS_{ABC}^2 + G_{12} c_1 c_2 MS_{AB} MS_{ABC} \\ &= (0.524)^2(1/6)^2(202.00)^2 + (1.3)^2(1/6)^2(13.84)^2 \\ &\quad + (-0.054)(1/6)(1/6)(202.00)(13.84) \\ &= 316.02 \end{aligned}$$

所以 $\sigma^2_{\tau\beta}$ 的近似 95%置信下限为

$$L = \hat{\sigma}^2_{\tau\beta} - \sqrt{\mathrm{V}_L} = 31.36 - \sqrt{316.02} = 13.58$$

该结果与 F 精确检验关于该效应的结果一致.

13.7.3 方差分量的最大似然估计

本章着重讲述关于方差分量估计的方差分析方法, 因为方差分析相对直接并且用的都是常见量, 即方差分析表中的那些均方. 不过, 该方法也有一些缺点, 包括有时会给出负的方差估计. 而且, 方差分析方法其实是一种**矩估计**(moment estimator) 的方法. 由于用矩估计得到的估计量的统计性质一般不太好, 数理统计学家一般不喜欢用矩估计. 数理统计学家更愿意使用的参数估计方法是**最大似然法**(method of maximum likelihood). 这种方法实施起来也许有些麻烦, 特别是对于实验设计模型. 但从某种意义来讲, 最大似然法选出的参数估计, 能针对具体模型与具体误差分布, 使样本观测结果出现的概率最大化. Milliken and Johnson(1984) 对最大似然法在实验设计模型中的应用给出了一个很好的综述.

最大似然法的完整介绍超出了本书的范围, 但它的基本思想还是容易说明的. 假设 x 是一个概率分布为 $f(x;\theta)$ 的随机变量, 其中 θ 是未知参数. 设 $x_1, x_2, \cdots, x_n$ 是有 n 个观测的随机样本, 则该样本的**似然函数**(likelihood function) 为

$$L(\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \cdot \cdots \cdot f(x_n;\theta)$$

注意, 似然函数只是一个关于未知参数 θ 的函数. θ 的最大似然估计是使似然函数 $L(\theta)$ 达到最大的 θ 的值.

为说明该方法如何用于带有随机效应的实验设计模型, 考虑 $a=b=n=2$ 的二因子模型. 此模型为

$$y_{ijk}=\mu+\tau_i+\beta_j+(\tau\beta)_{ij}+\varepsilon_{ijk}$$

其中 $i=1,2,\ j=1,2$ 且 $k=1,2$. 任一观测的方差是

$$\mathrm{V}(y_{ijk})=\sigma_y^2=\sigma_\tau^2+\sigma_\beta^2+\sigma_{\tau\beta}^2+\sigma^2$$

且协方差为

$$\mathrm{Cov}(y_{ijk},y_{i'j'k'})\begin{cases}=\sigma_\tau^2+\sigma_\beta^2+\sigma_{\tau\beta}^2 & i=i',j=j',k\neq k'\\ =\sigma_\tau^2 & i=i',j\neq j'\\ =\sigma_\beta^2 & i\neq i',j=j'\\ =0 & i\neq i',j\neq j'\end{cases}\tag{13.45}$$

不妨把观测值看作一个 8×1 向量, 记为

$$\boldsymbol{y}=\begin{bmatrix} y_{111}\\ y_{112}\\ y_{211}\\ y_{212}\\ y_{121}\\ y_{122}\\ y_{221}\\ y_{222}\end{bmatrix}$$

而方差与协方差可以表示为一个 8×8 的**协方差矩阵**:

$$\boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$$

其中 $\boldsymbol{\Sigma}_{11},\boldsymbol{\Sigma}_{22},\boldsymbol{\Sigma}_{12},\boldsymbol{\Sigma}_{21}=\boldsymbol{\Sigma}_{12}'$ 是如下的 4×4 的矩阵:

$$\boldsymbol{\Sigma}_{11}=\boldsymbol{\Sigma}_{22}=\begin{bmatrix}\sigma_y^2 & \sigma_\tau^2+\sigma_\beta^2+\sigma_{\tau\beta}^2 & \sigma_\tau^2 & \sigma_\tau^2\\ \sigma_\tau^2+\sigma_\beta^2+\sigma_{\tau\beta}^2 & \sigma_y^2 & \sigma_\tau^2 & \sigma_\tau^2\\ \sigma_\tau^2 & \sigma_\tau^2 & \sigma_y^2 & \sigma_\tau^2+\sigma_\beta^2+\sigma_{\tau\beta}^2\\ \sigma_\tau^2 & \sigma_\tau^2 & \sigma_\tau^2+\sigma_\beta^2+\sigma_{\tau\beta}^2 & \sigma_y^2\end{bmatrix}$$

$$\boldsymbol{\Sigma}_{12}=\begin{bmatrix}\sigma_\beta^2 & \sigma_\beta^2 & 0 & 0\\ \sigma_\beta^2 & \sigma_\beta^2 & 0 & 0\\ 0 & 0 & \sigma_\beta^2 & \sigma_\beta^2\\ 0 & 0 & \beta_\beta^2 & \sigma_\beta^2\end{bmatrix}$$

且 $\boldsymbol{\Sigma}_{21}$ 就是 $\boldsymbol{\Sigma}_{12}$ 的转置. 现每个观测都服从方差为 σ_y^2 的正态分布, 假定所有 $N=abn$ 个观测服从联合正态分布, 则随机模型的似然函数为

$$L(\mu,\sigma_\tau^2,\sigma_\beta^2,\sigma_{\tau\beta}^2,\sigma^2)=\frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{j}_N\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{j}_N\boldsymbol{\mu})\right]$$

其中 $\boldsymbol{j}_N$ 是 $N\times 1$ 的分量都为 1 的向量. $\mu,\sigma_\tau^2,\sigma_\beta^2,\sigma_{\tau\beta}^2,\sigma^2$ 的最大似然估计是使似然函数取到最大值时这些参数的取值. 另外, 还需要将方差分量的估计约束为非负值. 所以在实际中, 我们将在此约束下最大化似然函数.

由最大似然法估计方差分量的计算要用到特定的计算机软件. 一些通用统计软件包都具有此功能. SAS 用 SAS PROC MIXED 来计算随机或混合模型中的方差分量. 我们通过将 PROC MIXED 用于例 13.2 和 13.3 中的二因子析因模型, 来说明 PROC MIXED 的使用.

首先考虑例 13.2. 该模型是一个二因子析因的随机效应模型. 方差分析法得到了交互作用效应方差分量的一个负的估计值. 用 PROC MIXED 可以避免方差分量的负估计值, 只要指定使用**约束 (或残差) 最大似然估计方法**(REML). REML 实际上约束方差分量的估计为非负值.

SAS PROC MIXED 需要输入模型参数的协方差矩阵. 随机变量相互独立的随机模型的协方差结构为

$$\boldsymbol{G}=\begin{bmatrix}\sigma_\tau^2\boldsymbol{I} & 0 & 0\\ 0 & \sigma_\beta^2\boldsymbol{I} & 0\\ & 0 & \sigma_{\tau\beta}^2\boldsymbol{I}\end{bmatrix} \tag{13.46}$$

其中 $\boldsymbol{I}$ 是单位阵. 模型的该协差阵结构, 可以通过在 PROC MIXED 中 RANDOM 语句规定 TYPE=*structure* option 来实现. 例 13.2 中模型的协方差矩阵结构是用 TYPE=SIM (PROC MIXED 中的默认情形) 来规定的, 它规定了模型参数的协方差矩阵是 (13.46) 式给出的简单结构.

表 13.14 给出了用 SAS PROC MIXED 对例 13.2 中实验的分析结果. 我们选择了方差分量估计的 REML 方法. 输出的结果被标上了数字以便于下面进行解释.

(1) 方差分量估计与相关输出.

(2) 协方差参数. 类似于模型的参数: $\hat\sigma_\tau^2$,$\hat\sigma_\beta^2,\hat\sigma_{\tau\beta}^2,\hat\sigma^2$.

(3) 效应的方差估计与残差方差估计的比值: $\hat\sigma_i^2/\hat\sigma^2$.

(4) 参数估计. 这些是方差分量的 REML 估计值 $\hat\sigma_\tau^2,\hat\sigma_\beta^2,\hat\sigma_{\tau\beta}^2,\hat\sigma^2$. 要注意 $\sigma_{\tau\beta}^2$ 的 REML 估计值为 0.

(5) 估计的标准误. 这是参数估计的大样本标准误, $se(\hat\sigma_i^2)=\sqrt{\mathrm{V}(\hat\sigma_i^2)}$.

(6) 与方差估计有关的 Z 统计量: $Z=\hat\sigma_i^2/se(\hat\sigma_i^2)$.

(7) 算出的 Z 统计量的 P 值.

(8) 用于计算置信区间的 α 水平.

(9) 关于方差分量的 $100(1-\alpha)\%$ 的大样本正态理论置信区间的下限与上限:

$$L=\hat\sigma_i^2-Z_{\alpha/2}se(\hat\sigma_i^2),\quad U=\hat\sigma_i^2+Z_{\alpha/2}se(\hat\sigma_i^2)$$

(10) 估计的近似协方差矩阵. 这是方差分量估计的大样本协方差矩阵.

(11) 用于比较不同模型拟合程度的模型拟合度量.

注意, 用 SAS PROC MIXED 得到的结果, 与例 13.2 中用简化模型 (没有交互作用项) 拟合数据时所给出的结果相当一致.

在例 13.3 中, 考虑同样的实验, 但假定操作员是固定因子, 导出一个混合模型. SAS PROC

MIXED 能用于这种情况的方差分量的估计. 假定所有随机变量都是相互独立的 (即无约束混合模型), 观测值的协方差结构为

$$\operatorname{Cov}(y_{ijk}, y_{i'j'k'})\begin{cases} =\sigma_\beta^2+\sigma_{\tau\beta}^2+\sigma^2 & i=i', j=j', k=k' \\ =\sigma_\beta^2+\sigma_{\tau\beta}^2 & i=i', j=j', k\neq k' \\ =\sigma_\beta^2 & i\neq i', j=j' \\ =0 & j\neq j' \end{cases} \tag{13.47}$$

表 13.14 例 13.2 中量具可重复性与可再现性研究分析的 SAS PROC MIXED 输出 (使用方差分量的 REML 估计)

The MIXED Procedure

Class Level Information

Class	Levels	Values
PART	20	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
OPERATOR	3	1 2 3
REPLICAT	2	1 2

[1]

Covariance Parameter Estimates (REML)

[2] Cov Parm	[3] Ratio	[4] Estimate	[5] Std Errors	[6] Z	[7] Pr > |z|	[8] Alpha	[9] Lower	Upper
OPERATOR	0.01203539	0.01062922	0.03286000	0.32	0.7463	0.05	-0.0538	0.0750
PART	11.60743820	10.25126446	3.37376878	3.04	0.0024	0.05	3.6388	16.8637
PART*OPERATOR	-0.00000000	-0.00000000						
Residual	1.00000000	0.88316339	0.12616620	7.00	0.0000	0.05	0.6359	1.1304

[10]

Asymptotic Covariance Matrix of Estimates

Cov Parm	OPERATOR	PART	PART*OPERATOR	Residual
OPERATOR	0.00107978	0.00006632	0.00000000	-0.00039795
PART	0.00006632	11.38231579	0.00000000	-0.00265287
PART*OPERATOR	0.00000000	0.00000000	0.00000000	-0.00000000
Residual	-0.00039795	-0.00265287	-0.00000000	0.01591791

[11]

Model Fitting Information For VALUE

Description	Value
Observations	120.0000
Variance Estimate	0.8832
Standard Deviation Estimate	0.9398
REML Log Likelihood	-204.696
Akaike's Information Criterion	-208.696
Schwarz's Bayesian Criterion	-214.254
-2 REML Log Likelihood	409.3913

模型参数的协方差矩阵为

$$G=\begin{bmatrix}\sigma_\beta^2 I & 0\\ 0 & \sigma_{\tau\beta}^2 I\end{bmatrix} \tag{13.48}$$

(这是在调用 PROC MIXED 时通过选取 TYPE=SIM 规定的). SAS PROC MIXED 对例 13.3 的无约束混合模型的输出列在表 13.15 中. 这次仍然采用了 REML 方法. 对因子 "工件" 的方差估计与采用随机模型得到的估计非常相似. 另外, 输出中还包括了固定效应的 F 检验.

表 13.15 量具可重复性与可再现性研究分析的 SAS PROC MIXED 输出其中操作员作为固定效应, 使用方差分量的 REML 估计

The MIXED Procedure

Covariance Parameter Estimates (REML)

Cov Parm	Ratio	Estimate	Std Error	Z	PR > |z|	Alpha	Lower	Upper
PART	11.60743876	10.25126472	3.37376895	3.04	0.0024	0.05	3.6388	16.8637
PART*OPERATOR	0.00000000	0.00000000						
Residual	1.00000000	0.88316337	0.12616620	7.00	0.0000	0.05	0.6359	1.1304

Asymptotic Covariance Matrix of Estimates

Cov Parm	PART	PART*OPERATOR	Residual
PART	11.38231693	0.00000000	-0.00265287
PART*OPERATOR	0.00000000	0.00000000	-0.00000000
Residual	-0.00265287	0.00000000	-0.01591791

Model Fitting Information for VALUE

Description	Value
Observations	120.0000
Variance Estimate	0.8832
Standard Deviation Estimate	0.9398
REML Log Likelihodd	-204.729
AKaike's Information Criterion	-207.729
Schwarz's Bayesian Criterion	-211.872
-2 REML Log Likelihood	409.4572

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
OPERATOR	2	38	1.48	0.2401

13.8 思 考 题

*13.1 一家纺织厂有很多织布机. 假定每台织机每分钟产出的布相同. 为研究该假定, 随机选择 5 台织布机, 并记录它们在不同时间的产量. 所得数据如下:

织布机	产量 (1b/min)				
1	14.0	14.1	14.2	14.0	14.1
2	13.9	13.8	13.9	14.0	14.0
3	14.1	14.2	14.1	14.0	13.9
4	13.6	13.8	14.0	13.9	13.7
5	13.8	13.6	13.9	13.8	14.0

(a) 解释为什么这是一个随机效应实验. 各织机的产出是否相同?(取 $\alpha = 0.05$)

(b) 估计织机间的差异.

(c) 估计实验误差方差.

(d) 求 $\sigma_\tau^2/(\sigma_\tau^2+\sigma^2)$ 的 95%置信区间.

(e) 分析该试验的残差. 你认为该方差分析的假定条件是否满足?

13.2 一位制造商怀疑供货商提供的各批原料在含钙量上有显著差异. 目前仓库里有许多不同的批次. 从中随机选出 5 批进行研究. 一位化学家对每批做 5 次测定, 得到如下数据:

批次 1	批次 2	批次 3	批次 4	批次 5
23.46	23.59	23.51	23.28	23.29
23.48	23.46	23.64	23.40	23.46
23.56	23.42	23.46	23.37	23.37
23.39	23.49	23.52	23.46	23.32
23.40	23.50	23.49	23.39	23.38

(a) 各批次的含钙量有显著变化吗? (取 $\alpha = 0.05$)

(b) 估计方差分量.

(c) 求 $\sigma_\tau^2/(\sigma_\tau^2+\sigma^2)$ 的 95%置信区间.

(d) 分析该试验的残差. 方差分析的假定条件是否满足?

13.3 金属加工车间有几只炉子用于加热金属样品. 假设所有炉子都在同一温度上工作, 尽管这一假设不一定为真. 随机选出 3 只炉子, 测量它们的加热温度. 收集到的数据如下:

炉子	温　度					
1	491.50	498.30	498.10	493.50	493.60	
2	488.50	484.65	479.90	477.35		
3	490.10	484.80	488.25	473.00	471.85	478.65

(a) 炉子间的温度有显著差异吗? (取 $\alpha = 0.05$)

(b) 估计这一模型的方差分量.

(c) 分析该试验的残差, 并判断模型的适合性.

*13.4 *Journal of the Electrochemical Society* (Vol. 139, No. 2, 1992, pp. 524~532) 上的一篇论文提到一个研究多晶硅低压气相沉积的实验. 该实验是在得克萨斯州的奥斯汀的 Sematech 的大容量反应器中进行的. 该反应器有多个晶片位置, 从中随机选出 4 个. 响应变量是膜厚的均匀度. 实验进行 3 次重复, 数据如下:

晶片位置	均匀度		
1	2.76	5.67	4.49
2	1.43	1.70	2.19
3	2.34	1.97	1.47
4	0.94	1.36	1.65

(a) 晶片位置有显著差异吗? (取 $\alpha = 0.05$)

(b) 估计由于晶片位置造成的方差.

(c) 估计随机误差分量.

(d) 分析该试验的残差, 并评价模型的适合性.

13.5 考虑思考题 13.4 中的气相沉积实验.

(a) 估计均匀度响应的总变异.

(b) 均匀度响应的总变异中有多少是由反应器中位置的不同产生的?

(c) 若由反应器中不同位置所产生的变异可以消除, 则均匀度响应的变异可以降低到何种程度? 你认为是这种下降显著吗?

13.6 *Journal of Quality Technology* (Vol. 13, No. 2, 1981, pp. 111~114) 上的一篇论文提到一个研究 4 种化学漂白剂对纸浆白度影响的实验. 这 4 种化学漂白剂是从大量的漂白剂中随机地选取的. 数据如下:

漂白剂	纸浆白度				
1	77.199	74.466	92.746	76.208	82.876
2	80.522	79.306	81.914	80.346	73.385
3	79.417	78.017	91.596	80.802	80.626
4	78.001	78.358	77.544	77.364	77.386

(a) 漂白剂不同类型间有显著变化吗? (取 $\alpha = 0.05$)

(b) 估计由漂白剂类型产生的变异.

(c) 估计由随机误差产生的变异.

(d) 分析试验的残差, 并评价模型的适合性.

13.7 考虑单因子平衡随机效应方法. 试找出一个求 $\sigma^2/(\sigma_\tau^2 + \sigma^2)$ 的 $100(1-\alpha)\%$ 的置信区间的方法.

13.8 考虑思考题 13.1.

(a) 若 σ_τ^2 是误差方差 σ^2 的 4 倍, 则接受假设 H_0 的概率是多少?

(b) 若织布机之间的差异大得足以使观测值的标准差增加 20%, 我们要至少以 0.80 的概率检测出这一变异来源. 应使用多大的样本量?

*13.9 进行一项实验, 来研究一个测量系统的能力. 随机选取 10 个工件, 并且随机选出两位的操作员各自将它们重复测量 3 次. 测量次序是随机的, 数据如下:

工件编号	操作员 1 测量值			操作员 2 测量值		
	1	2	3	1	2	3
1	50	49	50	50	48	51
2	52	52	51	51	51	51
3	53	50	50	54	52	51
4	49	51	50	48	50	51
5	48	49	48	48	49	48
6	52	50	50	52	50	50
7	51	51	51	51	50	50
8	52	50	49	53	48	50
9	50	51	50	51	48	49
10	47	46	49	46	47	48

*13.10 Hoof 和 Berman 的一篇论文 ("Statistical Analysis of Power Module Thermal Test Equipment

Performance," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology* Vol. 11, pp. 516~520, 1988) 描述了这样一个实验, 该实验研究一种用于感应电机启动器的功率模块的热抗能力 (°C/w×100). 有 10 个部件, 3 位检验员, 重复 3 次. 数据如下表所示:

工件编号	检验员 1 检验值 1	2	3	检验员 2 检验值 1	2	3	检验员 3 检验值 1	2	3
1	37	38	37	41	41	40	41	42	41
2	42	41	43	42	42	42	43	42	43
3	30	31	31	31	31	31	29	30	28
4	42	43	42	43	43	43	42	42	42
5	28	30	29	29	30	29	31	29	29
6	42	42	43	45	45	45	44	46	45
7	25	26	27	28	28	30	29	27	27
8	40	40	40	43	42	42	43	43	41
9	25	25	25	27	29	28	26	26	26
10	35	34	34	35	35	34	35	34	35

(a) 假定工件与检验员都是随机效应, 分析该实验的数据.

(b) 用方差分析方法估计方差分量.

*13.11 再考虑思考题 5.6 中的数据. 假设两个因子 (机器与操作员) 都是随机选择的.

(a) 分析该实验的数据.

(b) 用方差分析方法计算方差分量的点估计.

13.12 再考虑思考题 5.13 中的数据. 假定两个因子都是随机的.

(a) 分析该实验的数据.

(b) 估计方差分量.

13.13 假设思考题 5.11 中炉的位置是随机选择的, 从而是一个混合模型实验. 在此新假定下重新分析该实验的数据. 估计合适的模型分量.

*13.14 重新分析思考题 13.9 中的测量系统实验, 假定操作员是固定因子. 估计合适的模型分量.

13.15 重新分析思考题 13.10 中的测量系统实验, 假定检验员是固定因子. 估计合适的模型分量.

13.16 思考题 5.6 中, 假设只对 4 台机器感兴趣, 但操作员是随机选择的.

(a) 哪种模型是适合的?

(b) 进行分析并估计模型分量.

13.17 采用约束模型的假定, 通过计算期望, 求出二因子析因设计的混合模型的各均方的期望值. 将你的结果与 (13.23) 式中给出的期望均方进行比较, 看看是否一样.

13.18 考虑例 13.6 中的三因子设计. 给所有的主效应和交互效应提出合适的检验统计量. 对 A, B 固定且 C 随机的情况重做一次.

13.19 考虑例 13.7 中的实验. 对 A, B, C 都是随机的情况进行分析.

13.20 推导表 13.11 中的期望均方.

13.21 考虑四因子析因实验, 其中因子 A 有 a 个水平, 因子 B 有 b 个水平, 因子 C 有 c 个水平, 因子 D 有 d 个水平, 有 n 次重复. 对下列情况求出平方和、自由度以及期望均方. 下面的所有模型都假定是约束模型. 可以使用诸如 Minitab 之类的计算软件.

(a) A, B, C, D 都是固定因子.

(b) A, B, C, D 都是随机因子.

(c) A 固定, B, C, D 都是随机的.

(d) A, B 固定, 且 C 与 D 随机.

(e) A, B, C 固定, D 随机.

所有的效应存在精确的检验法吗? 如果不是, 给那些不能直接检验的效应提出检验统计量.

13.22 重新考虑 13.21 中的情况 (c), (d), (e). 假定是无约束模型, 求出期望均方. 可以使用诸如 Minitab 之类的计算机软件. 将这些结果与约束模型的结果进行比较.

13.23 在思考题 5.17 中, 设 3 名操作工是随机选取的. 在这些条件下分析那些数据并写出结论. 估计方差分量.

13.24 考虑三因子析因模型

$$y_{ijk} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\beta\gamma)_{jk} + \varepsilon_{ijk} \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b \\ k = 1, 2, \cdots, c \end{cases}$$

假定所有因子都是随机的, 写出包括期望均方的方差分析表. 给所有的效应提出恰当的检验统计量.

13.25 只有一次重复的三因子析因模型是

$$y_{ijk} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\beta\gamma)_{jk} + (\tau\gamma)_{ik} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijk}$$

如果所有因子都是随机的. 任一效应都能检验吗? 如果不存在三因子交互作用以及交互作用 $(\tau\beta)_{ij}$, 能检验所有其余的效应吗?

13.26 在思考题 5.6 中, 假设机器与操作员都是随机选择的. 确定用于识别机器效应 $\sigma_\beta^2 = \sigma^2$ 这一检验的势, 其中 σ_β^2 是机器因子的方差分量. 重复两次是否足够?

13.27 在二因子混合模型的方差分析中, 证明 $\mathrm{Cov}[(\tau\beta)_{ij}, (\tau\beta)_{i'j}] = -(1/a)\sigma_{\tau\beta}^2$, $i \neq i'$.

13.28 证明在任意随机或固定模型中, 方差分析法总能得到各方差分量的无偏估计.

13.29 采用通常的正态性假定, 推导出用方差分析法会得出负的方差分量的概率的表示式. 由此结果, 写出单因子方差分析中得到 $\sigma_\tau^2 < 0$ 的概率. 谈谈该概率结果的意义.

13.30 分别采用约束与无约束混合模型, 分析思考题 13.9 中的数据. 比较两种模型下的结果.

13.31 考虑二因子混合模型, 证明固定因子均值 (如 A) 的标准误为 $[MS_{AB}/bn]^{1/2}$.

13.32 考虑思考题 13.9 中随机模型的方差分量.

(a) 求 σ^2 的精确 95% 置信区间.

(b) 用 Satterthwaite 法求其他方差分量的近似 95% 置信区间.

13.33 对思考题 5.6 中描述的实验, 假定两个因子都是随机的. 求 σ^2 的精确 95% 置信区间. 用 Satterthwaite 法构造其他方差分量的近似 95% 置信区间.

13.34 考虑思考题 5.17 中的三因子实验, 并假定操作工是随机选择的. 求操作工方差分量的近似 95% 置信区间.

13.35 用 13.7.2 节介绍的修正的大样本方法重做思考题 13.30. 比较并讨论两类置信区间的差异.

13.36 用 13.7.2 节介绍的修正的大样本方法重做思考题 13.32. 比较并讨论这种置信区间与原来得到的置信区间的差异.

第 14 章　嵌套设计和裂区设计

本章纲要

14.1　二级嵌套设计
14.1.1　统计分析
14.1.2　诊断检测
14.1.3　方差分量
14.1.4　交错嵌套设计
14.2　一般的 m 级嵌套设计
14.3　含被套因子和交叉因子的设计
14.4　裂区设计
14.5　裂区设计的其他变形
14.5.1　多于两个因子的裂区设计
14.5.2　裂裂区设计
14.5.3　带裂区设计
第 14 章补充材料
S14.1　交错嵌套设计
S14.2　疏忽大意的裂区设计

本章介绍两类重要的实验设计：**嵌套设计** (nested design) 和**裂区设计** (split-plot design). 这两类设计在工业实验设计中有着相当广泛的应用. 它们常常含有一个或多个**随机因子**, 所以这里将用到第 13 章中介绍的一些概念.

14.1　二级嵌套设计

在某些多因子实验中, 一个因子 (例如, 因子 B) 的各个水平相对于另一个因子 (例如, 因子 A) 的各个水平而言, 相似但不相同. 这样的设计叫做以因子 B 的水平被套在因子 A 的水平下的**嵌套设计或分级设计**. 例如, 考虑一家公司从 3 个不同的货主手中购买原料. 公司想确定各个货主供应的原料的纯度是否相同. 从每位货主可得 4 批原料, 对每批原料要进行 3 次纯度测量. 这种情况画在图 14.1 中.

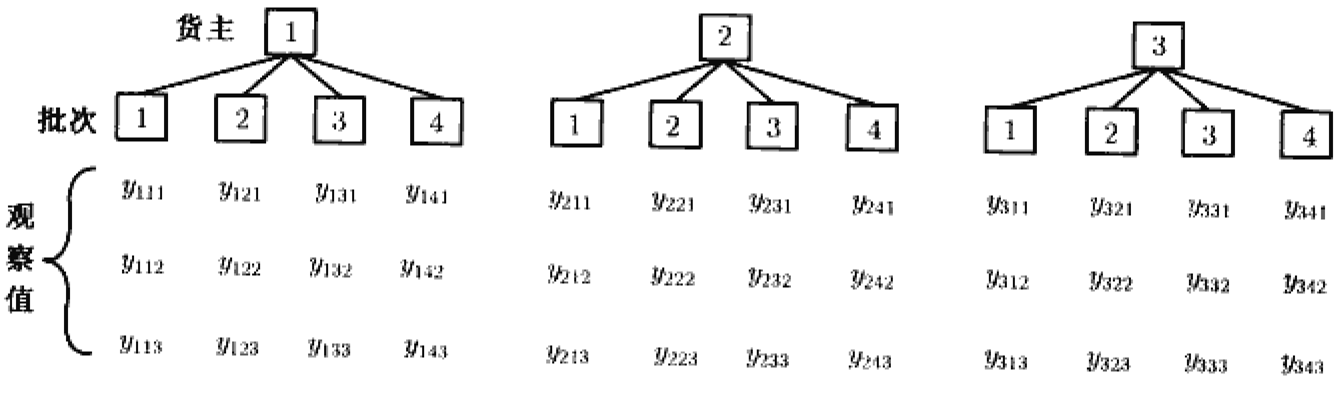


图 14.1　二级嵌套设计

这是一个**二级嵌套设计**, 批次被套在货主下面. 乍一看, 你可能会问, 为什么这不是一个析因实验呢? 若这是析因实验, 则批次 1 总是归属于同一批, 批次 2 总是归属于同一批, 等等. 现在, 显然不是这种情况, 因为每位货主供应的批次**只属于**那位特定的货主. 也就是说, 货主 1 供应的批次 1 和其他任何一位货主供应的批次 1 没有关系, 货主 1 供应的批次 2 和其他任何一位货主供应的批次 2 没有关系, 等等. 为了强调每位货主供应的批次不同这一事实, 也可以重新编

号为 1, 2, 3, 4 是货主 1 供应的; 5, 6, 7, 8 是货主 2 供应的; 9, 10, 11, 12 是货主 3 供应的, 如图 14.2 所示.

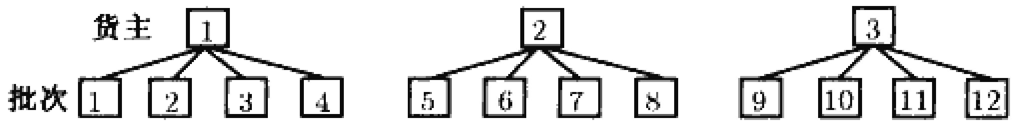


图 14.2 二级嵌套设计的另一编排

有时, 我们也许不能确定一个因子究竟是被交叉在一个析因设计中还是被嵌套在一个析因设计中. 如果因子的各个水平可以任意地如图 14.2 那样重新编号, 则这因子就是被嵌套的.

14.1.1 统计分析

二级嵌套设计的线性统计模型是

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k} \quad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b \\ k = 1, 2, \cdots, n \end{cases} \tag{14.1}$$

也就是说, 因子 A 有 a 个水平, 因子 B 的 b 个水平被套在 A 的每个水平下, n 是重复数. 下标 $j(i)$ 表示因子 B 的第 j 个水平被嵌套在因子 A 的第 i 个水平下. 把重复试验看作为被嵌套在 A 和 B 的水平组合之内是方便的, 这样一来, 下标 $(ij)k$ 表示误差项. 因为在 A 的每个水平内 B 的水平数相等以及重复数相等, 所以这一设计是**平衡嵌套设计**. 因为因子 B 的每个水平不能和因子 A 的每个水平一起出现, 所以 A 和 B 之间没有交互作用.

可以把总的校正平方和写为

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{\cdots})^2=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}[(\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdots})+(\bar{y}_{ij\cdot}-\bar{y}_{i\cdot\cdot})+(y_{ijk}-\bar{y}_{ij\cdot})]^2 \tag{14.2}$$

展开 (14.2) 式的右边得

$$\begin{aligned}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{\cdots})^2 &= bn\sum_{i=1}^{a}(\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdots})^2 \\ &\quad + n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij\cdot}-\bar{y}_{i\cdot\cdot})^2+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{ij\cdot})^2\end{aligned} \tag{14.3}$$

因为 3 个交叉乘积项为零. (14.3) 式表明, 总的平方和可以分解为因子 A 的平方和、在 A 的水平下因子 B 的平方和以及误差的平方和. 用符号表示, 可以把 (14.3) 式写为

$$SS_T = SS_A + SS_{B(A)} + SS_E \tag{14.4}$$

其中 SS_T 有 $abn-1$ 个自由度, SS_A 有 $a-1$ 个自由度, $SS_{B(A)}$ 有 $a(b-1)$ 个自由度, 误差的自由度为 $ab(n-1)$. 注意, $abn-1=(a-1)+a(b-1)+ab(n-1)$. 如果误差是 $\mathrm{NID}(0,\sigma^2)$, 我们就可以将 (14.4) 式右边每项平方和除以它的自由度来求得独立分布的均方, 使得任意两个均方的比服从 F 分布.

用来检验因子 A 和因子 B 效应的恰当统计量依赖于 A 和 B 究竟是**固定的还是随机的**. 如果因子 A 和 B 是固定的, 我们假定 $\sum\limits_{i=1}^{a}\tau_i=0$ 和 $\sum\limits_{j=1}^{b}\beta_{j(i)}=0(i=1,2,\cdots,a)$. 也就是说, A

的处理效应和为零，在 A 的每个水平内，B 的处理效应和为零. 如果 A 和 B 是随机的，则可以假定 τ_i 是 $\text{NID}(0,\sigma_\tau^2)$ 以及 $\beta_{j(i)}$ 是 $\text{NID}(0,\sigma_\beta^2)$. A 固定 B 随机的混合模型也很常见. 期望均方可以直接用第 13 章的法则来确定. 表 14.1 给出了这些情况的期望均方.

表 14.1 二级嵌套设计的期望均方

$\text{E}(MS)$	A 固定，B 固定	A 固定，B 随机	A 随机，B 随机
$\text{E}(MS_A)$	$\sigma^2+\frac{bn\sum\tau_i^2}{a-1}$	$\sigma^2+n\sigma_\beta^2+\frac{bn\sum\tau_i^2}{a-1}$	$\sigma^2+n\sigma_\beta^2+bn\sigma_\tau^2$
$\text{E}(MS_{B(A)})$	$\sigma^2+\frac{n\sum\sum\beta_{j(i)}^2}{a(b-1)}$	$\sigma^2+n\sigma_\beta^2$	$\sigma^2+n\sigma_\beta^2$
$\text{E}(MS_E)$	σ^2	σ^2	σ^2

表 14.1 表明，如果 A 和 B 的水平是固定的，$H_0:\tau_i=0$ 用 MS_A/MS_E 检验，而 $H_0:\beta_{j(i)}=0$ 则用 $MS_{B(A)}/MS_E$ 检验. 如果 A 是固定因子，B 是随机因子，则 $H_0:\tau_i=0$ 用 $MS_A/MS_{B(A)}$ 检验，而 $H_0:\sigma_\beta^2=0$ 用 $MS_{B(A)}/MS_E$ 检验. 最后，如果 A 和 B 都是随机因子，则检验 $H_0:\sigma_\tau^2=0$ 用 $MS_A/MS_{B(A)}$，检验 $H_0:\sigma_\beta^2=0$ 用 $MS_{B(A)}/MS_E$. 检验方法概括在表 14.2 所示的方差分析表中. 这些平方和的计算公式通过展开 (14.3) 式的量并加以简化就可求得. 它们是

$$SS_A=\frac{1}{bn}\sum_{i=1}^{a}y_{i..}^2-\frac{y_{...}^2}{abn} \tag{14.5}$$

$$SS_{B(A)}=\frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij.}^2-\frac{1}{bn}\sum_{i=1}^{a}y_{i..}^2 \tag{14.6}$$

$$SS_E=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}^2-\frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij.}^2 \tag{14.7}$$

$$SS_T=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}^2-\frac{y_{...}^2}{abn} \tag{14.8}$$

表 14.2 二级嵌套设计的方差分析表

方差来源	平方和	自由度	均方
A	$bn\sum(\bar{y}_{i..}-\bar{y}_{...})^2$	$a-1$	MS_A
A 内的 B	$n\sum\sum(\bar{y}_{ij.}-\bar{y}_{i..})^2$	$a(b-1)$	$MS_{B(A)}$
误差	$\sum\sum\sum(y_{ijk}-\bar{y}_{ij.})^2$	$ab(n-1)$	MS_E
总和	$\sum\sum\sum(y_{ijk}-\bar{y}_{...})^2$	$abn-1$	

计算 $SS_{B(A)}$ 的 (14.6) 式可写为

$$SS_{B(A)}=\sum_{i=1}^{a}\left[\frac{1}{n}\sum_{j=1}^{b}y_{ij.}^2-\frac{y_{i..}^2}{bn}\right]$$

这个计算公式说明：要求 $SS_{B(A)}$，只要先求出在 A 的每个水平下 B 的水平间的平方和，然后再对 A 的所有水平求和即可.

例 14.1 考虑一家公司, 它从 3 位不同的货主中买了几批原料. 原料的纯度差别相当大, 这在制造成品时会引起问题. 我们想要确定, 纯度的差异性是否归因于货主之间的差异. 从每位货主中随机选取 4 批原料, 每批进行 3 次纯度测量. 当然, 这是一个二级嵌套设计. 将测量数据减去 93 后得到的规范数据如表 14.3 所示.

表 14.3 例 14.1 的纯度规范数据 (规范值: y_{ijk}=纯度 −93)

		货主 1				货主 2				货主 3			
	批次	1	2	3	4	1	2	3	4	1	2	3	4
		1	−2	−2	1	1	0	−1	0	2	−2	1	3
		−1	−3	0	4	−2	4	0	3	4	0	−1	2
		0	−4	1	0	−3	2	−2	2	0	2	2	1
批次总和	$y_{ij.}$	0	−9	−1	5	−4	6	−3	5	6	0	2	6
货主总和	$y_{i..}$	−5				4				14			

算得的平方和如下:

$$SS_T=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}^2-\frac{y_{...}^2}{abn}=153.00-\frac{(13)^2}{36}=148.31$$

$$SS_A=\frac{1}{bn}\sum_{i=1}^{a}y_{i..}^2-\frac{y_{...}^2}{abn}=\frac{1}{(4)(3)}[(-5)^2+(4)^2+(14)^2]-\frac{(13)^2}{36}=15.06$$

$$SS_{B(A)}=\frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij.}^2-\frac{1}{bn}\sum_{i=1}^{a}y_{i..}^2$$

$$=\frac{1}{3}[(0)^2+(-9)^2+(-1)^2+\dots+(2)^2+(6)^2]-19.75=69.92$$

$$SS_E=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}^2-\frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij.}^2=153.00-89.67=63.33$$

方差分析见表 14.4. 货主是固定的, 批次是随机的, 所以期望均方可由表 14.1 中间一列求得. 为方便起见, 在表 14.4 中重写这一列. 通过查看 P 值知道, 货主对纯度没有显著性的效应, 但是, 同一货主供应的各批次原料间的纯度有显著性差异.

表 14.4 例 14.1 的数据的方差分析

方差来源	平方和	自由度	均方	期望均方	均值 F_0	P 值
货主	15.06	2	7.53	$\sigma^2+3\sigma_\beta^2+6\sum\tau_i^2$	0.97	0.42
批次 (货主内的)	69.92	9	7.77	$\sigma^2+3\sigma_\beta^2$	2.94	0.02
误差	63.33	24	2.64	σ^2		
总和	148.31	35				

这一实验的实际意义和分析十分重要. 实验者的目的是寻找原料纯度变异性的来源. 如果它是来自货主间的差异, 则可以选择 “最好的” 货主来解决此问题. 但是, 这种解决问题的办法在此处不适用, 因为变异性的主要来源是货主内各批原料纯度之间的差异. 因此, 我们要解决这个问题, 就必须和货主一起努力来减少他们各批次之间的变异. 这就可能涉及改进货主的生产过程或他们内部的质量保证系统.

如果我们错误地将这一设计作为二因子析因实验来分析，将会发生什么情况呢？如果认为批次与货主是交叉的，对有 3 次重复的每一批次 × 货主单元求得批次的总和是 2, −3, −2, 16. 这样一来，就可算出批次的平方和与交互作用的平方和. 假定是混合模型，完全析因的方差分析如表 14.5 所示.

表 14.5 例 14.1 的二级嵌套设计作为析因设计的错误分析 (货主固定，批次随机)

方差来源	平方和	自由度	期望均方	F_0	P 值
货主 (S)	15.06	2	7.53	1.02	0.42
批次 (B)	25.64	3	8.55	3.24	0.04
$S \times B$ 交互作用	44.28	6	7.38	2.80	0.03
误差	63.33	24	2.64		
总和	148.31	35			

此分析表明，批次有显著性差异，且批次与货主间的交互作用也是显著的. 但是，批次 × 货主交互作用难于给出实际的解释. 例如，这一显著性效应意味着货主效应在不同的批次间是变化的吗？还有，这一显著性的交互作用与不显著的货主效应一起会导致分析者得出这样的结论，即货主实际上是有差异的，但它们的效应被显著性的交互作用掩饰起来了.

计算

一些统计软件包可以处理嵌套设计的分析. 表 14.6 给出了 Minitab 中 (使用约束模型) 的平衡 ANOVA 过程的输出结果，这个数值结果与表 14.4 的手工计算结果一致. Minitab 还在表 14.6 的下半部分给出了期望均方. 注意符号 $Q[1]$ 是代表货主的固定效应的平方项，用我们的记号是：

$$Q[1] = \frac{\sum_{i=1}^{a} \tau_i^2}{a-1}$$

表 14.6 例 14.1 的 Minitab (Balanced ANOVA) 输出结果

```
Analysis of Variance (Balanced Designs)

Factor              Type    Levels   values
Supplier            fixed        3        1        2        3
Batch (Supplier)    random       4        1        2        3        4

Analysis of Variance for Purity

Source             DF        SS        MS       F        P
Supplier            2    15.056     7.528    0.97    0.416
Batch (Supplier)    9    69.917     7.769    2.94    0.017
Error              24    63.333     2.639
Total              35   148.306

Source                Variance  Error  Expected Mean Square for Each Term
                      component  term  (using restricted model)
1 Supplier                          2  (3) + 3(2) + 12Q[1]
2 Batch (Supplier)       1.710      3  (3) + 3(2)
3 Error                  2.639         (3)
```

因此, 用 Minitab 中的固定效应项表示货主的期望均方是 $12Q[1]=12\sum_{i=1}^{3}\tau_i^2/(3-1)=6\sum_{i=1}^{3}\tau_i^2$, 这与表 14.4 中的列表算法是一致的.

有时候手头没有特别为嵌套设计编的计算机程序, 则注意比较表 14.4 和表 14.5 可知:

$$SS_B + SS_{S\times B} = 25.64 + 44.28 = 69.92 \equiv SS_{B(S)}$$

也就是说, 货主内批次的平方和由批次的平方和与批次 × 货主交互作用的平方和所组成. 其自由度有类似的性质, 也就是说,

$$\frac{\text{批次}}{3} + \frac{\text{批次}\times\text{货主}}{6} = \frac{\text{货主内的批次}}{9}$$

因此, 分式析因设计的计算机程序也可用来分析嵌套设计, 只要把被嵌套因子的 "主效应" 与被嵌套因子和嵌套因子的交互作用合并起来即可.

14.1.2 诊断检测

诊断检测所用的主要手段是**残差分析**. 二级嵌套设计的残差是

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk}$$

而拟合值是

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_{j(i)}$$

如果对模型参数加上通常的限制 ($\sum_i \hat{\tau}_i = 0$ 和 $\sum_j \hat{\beta}_{j(i)} = 0,\ i=1,2,\cdots,a$), 则 $\hat{\mu}=\bar{y}_{...}$, $\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}$, $\hat{\beta}_{j(i)} = \bar{y}_{ij.} - \bar{y}_{i..}$. 于是拟合值是

$$\hat{y}_{ijk} = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..}) = \bar{y}_{ij.}$$

这样一来, 二级嵌套设计的残差是

$$e_{ijk} = y_{ijk} - \bar{y}_{ij.} \tag{14.9}$$

其中 $\bar{y}_{ij.}$ 是各批次平均值.

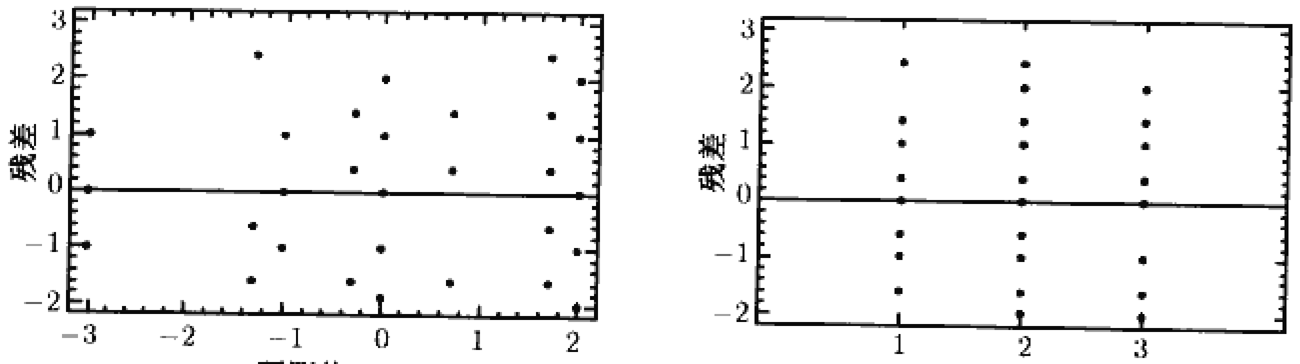
例 14.1 中纯度的观测值、拟合值以及残差的数据如下表.

观察值 y_{ijk}	拟合值 $\hat{y}_{ijk}=\bar{y}_{ij.}$	$e_{ijk}=y_{ijk}-\bar{y}_{ij.}$	观察值 y_{ijk}	拟合值 $\hat{y}_{ijk}=\bar{y}_{ij.}$	$e_{ijk}=y_{ijk}-\bar{y}_{ij.}$
1	0.00	1.00	4	1.67	2.33
−1	0.00	−1.00	0	1.67	−1.67
0	0.00	0.00	1	−1.33	2.33
−2	−3.00	1.00	−2	−1.33	−0.67
−3	−3.00	0.00	−3	−1.33	−1.67
−4	−3.00	−1.00	0	2.00	−2.00
−2	−0.33	−1.67	4	2.00	2.00
0	−0.33	0.33	2	2.00	0.00
1	−0.33	1.33	−1	−1.00	0.00
1	1.67	−0.67	0	−1.00	1.00

(续)

观察值 y_{ijk}	拟合值 $\hat{y}_{ijk}=\bar{y}_{ij.}$	$e_{ijk}=y_{ijk}-\bar{y}_{ij.}$	观察值 y_{ijk}	拟合值 $\hat{y}_{ijk}=\bar{y}_{ij.}$	$e_{ijk}=y_{ijk}-\bar{y}_{ij.}$
−2	−1.00	−1.00	0	0.00	0.00
0	1.67	−1.67	2	0.00	2.00
3	1.67	1.33	1	0.67	0.33
2	1.67	0.33	−1	0.67	−1.67
2	2.00	0.00	2	0.67	1.33
4	2.00	2.00	3	2.00	1.00
0	2.00	−2.00	2	2.00	0.00
−2	0.00	−2.00	1	2.00	−1.00

通常的诊断检测 —— 包括正态概率图、检测离群值和画出拟合值与残差的关系图 —— 现在都可进行了. 作为说明, 图 14.3 画出了残差与拟合值的关系图以及残差与货主因子水平的关系图.



(a) 预测值与残差的关系图

(b) 货主与残差的关系图

图 14.3 例 14.1 的残差图

在例 14.1 所描述的那种情形下, 残差图因其包含的附加诊断信息而特别有用. 例如, 方差分析显示所有货主供应的原料的纯度平均值没有显著差异, 但批次之间的差异有统计上的显著性 (即 $\sigma_\beta^2>0$). 但是所有货主的批内差异都相同吗? 事实上我们假设了它们都相同, 如果这个假设不成立, 我们当然也想了解它, 因为这对实验结果的解释有着相当大的实际影响. 图 14.3b 给出了残差与货主的关系图, 它很简单, 但对这个假设的检验很有效. 因为所有 3 个货主的残差分布大致相同, 所以我们可以认为, 所有这 3 个货主的批次之间的纯度差异大致一样.

14.1.3 方差分量

对随机效应情形, 方差分析法可用来估计方差分量 σ^2, σ_β^2, σ_τ^2. 由表 14.1 最后一列的期望均方, 得

$$\hat{\sigma}^2=MS_E \tag{14.10}$$

$$\hat{\sigma}_\beta^2=\frac{MS_{B(A)}-MS_E}{n} \tag{14.11}$$

$$\hat{\sigma}_\tau^2=\frac{MS_A-MS_{B(A)}}{bn} \tag{14.12}$$

很多嵌套设计的应用涉及主因子 (A) 固定、被嵌套因子 (B) 随机的**混合模型**. 这就是例 14.1 所描述的情形; 货主 (因子 A) 是固定的, 而原料的批次 (因子 B) 是随机的. 货主的效应可估计为

$$\hat{\tau}_1 = \bar{y}_{1..} - \bar{y}_{...} = \frac{-5}{12} - \frac{13}{36} = \frac{-28}{36}$$
$$\hat{\tau}_2 = \bar{y}_{2..} - \bar{y}_{...} = \frac{4}{12} - \frac{13}{36} = \frac{-1}{36}$$
$$\hat{\tau}_3 = \bar{y}_{3..} - \bar{y}_{...} = \frac{4}{12} - \frac{13}{36} = \frac{29}{36}$$

为估计方差分量 σ^2 和 σ_β^2, 我们删去方差分析表中属于货主的那一行, 并应用方差分析估计法于另外两行, 得

$$\hat{\sigma}^2 = MS_E = 2.64$$
$$\hat{\sigma}_\beta^2 = \frac{MS_{B(A)} - MS_E}{n} = \frac{7.77 - 2.64}{3} = 1.71$$

这些结果也在表 14.6 中 Minitab 输出的下半部分给出. 从例 14.1 的分析, 我们知道: τ_i 与零没有显著性差异, 而方差分量 σ_β^2 则远大于零.

14.1.4 交错嵌套设计

嵌套设计的应用中可能会碰到这种问题, 有时为了在最上层处得到数目比较合理的自由度, 我们可以在较低层去掉很多 (也许非常多的) 自由度. 例如, 假设我们对多份不同材料研究化学分析的差异, 我们计划对每份材料取 5 个样品, 每个样品测量两次. 如果想要估计材料的方差分量, 不妨从中选择 10 份不同的材料, 于是材料有 9 个自由度, 样品有 40 个自由度, 测量有 50 个自由度.

避免这样做的一种方法是使用一种特殊的不平衡嵌套设计, 称之为**交错嵌套设计**(staggered nested design). 图 14.4 给出了交错嵌套设计的一个例子. 注意, 每份材料只取两个样品, 其中一个样品被测量两次, 而另一个样品只被测量一次. 如果共有 a 份材料, 则材料 (即一般情况下的上层) 有 $a-1$ 个自由度, 且所有下层都恰有 a 个自由度. 关于这种设计的更多使用和分析, 参见 Bainbridge (1965), Smith and Beverly (1981), 以及 Nelson (1983,1995a,1995b). 本章的**补充阅读材料**中给出了一个交错嵌套设计的完整例子.

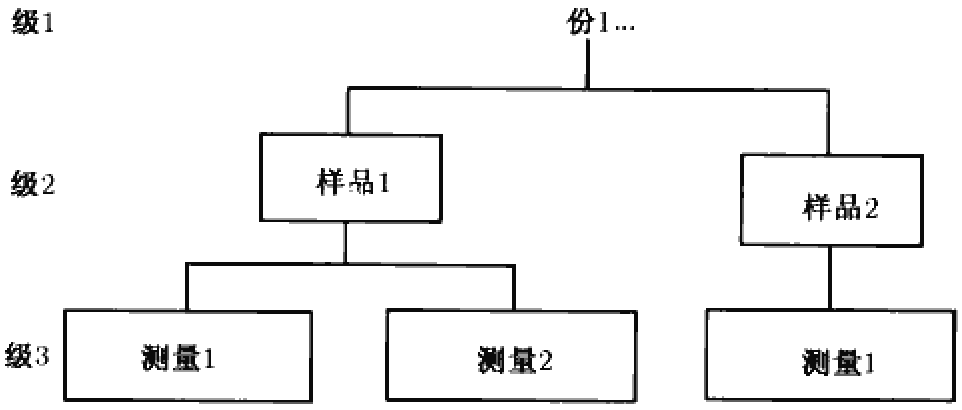


图 14.4 一个三级交错嵌套设计

14.2 一般的 m 级嵌套设计

14.1 节的结果容易推广到 m 个完全被套因子的情形. 这种设计叫做 **m 级嵌套设计**. 作为一个例子, 设铸造厂要研究两种不同配方合金的硬度. 每种合金配方各炼 3 炉, 从每炉中随机选取两块铸件, 从每块铸件上测量硬度两次. 图 14.5 说明了这种情况.

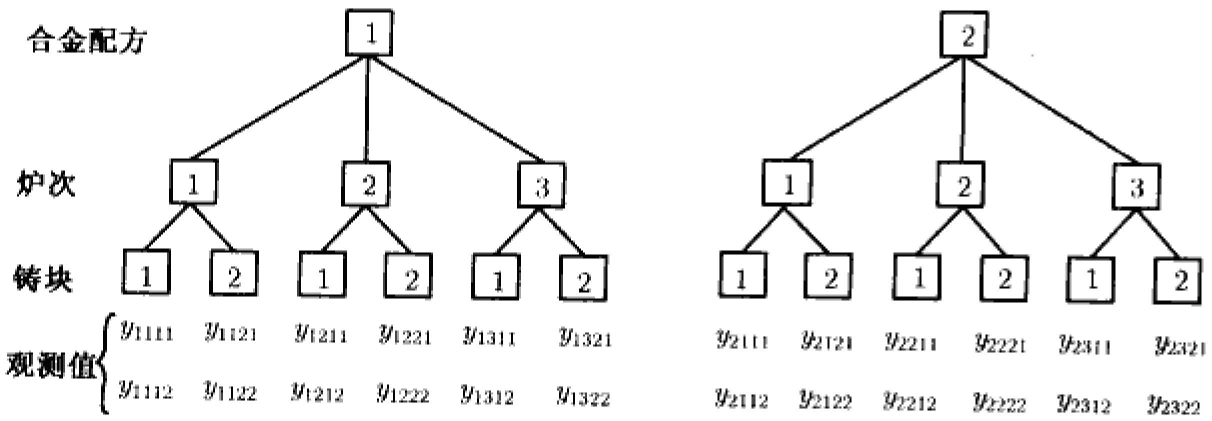


图 14.5　三级嵌套设计

在此实验中, 炉次被嵌套在合金配方因子的水平下, 铸块被嵌套在炉次因子的水平下. 因此, 这是有两次重复的三级嵌套设计.

一般三级嵌套设计的模型是

$$y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_{k(ij)} + \varepsilon_{(ijk)l} \quad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b \\ k = 1, 2, \cdots, c \\ l = 1, 2, \cdots, n \end{cases} \tag{14.13}$$

就我们的例子来说, τ_i 是第 i 种合金配方的效应, $\beta_{j(i)}$ 是第 i 种合金的第 j 炉的效应, $\gamma_{k(ij)}$ 是第 k 块铸块在第 i 种合金的第 j 炉下的效应, $\varepsilon_{(ijk)l}$ 是通常的 NID(0, σ^2) 误差项. 这一模型可直接推广至 m 个因子的情形.

在上面的例子中, 硬度的总变异性由 3 个分量组成: 一个由合金配方所得, 一个由炉次所得, 另一个由分析测试误差所得. 硬度总变异性的这些分量可用图 14.6 来说明.

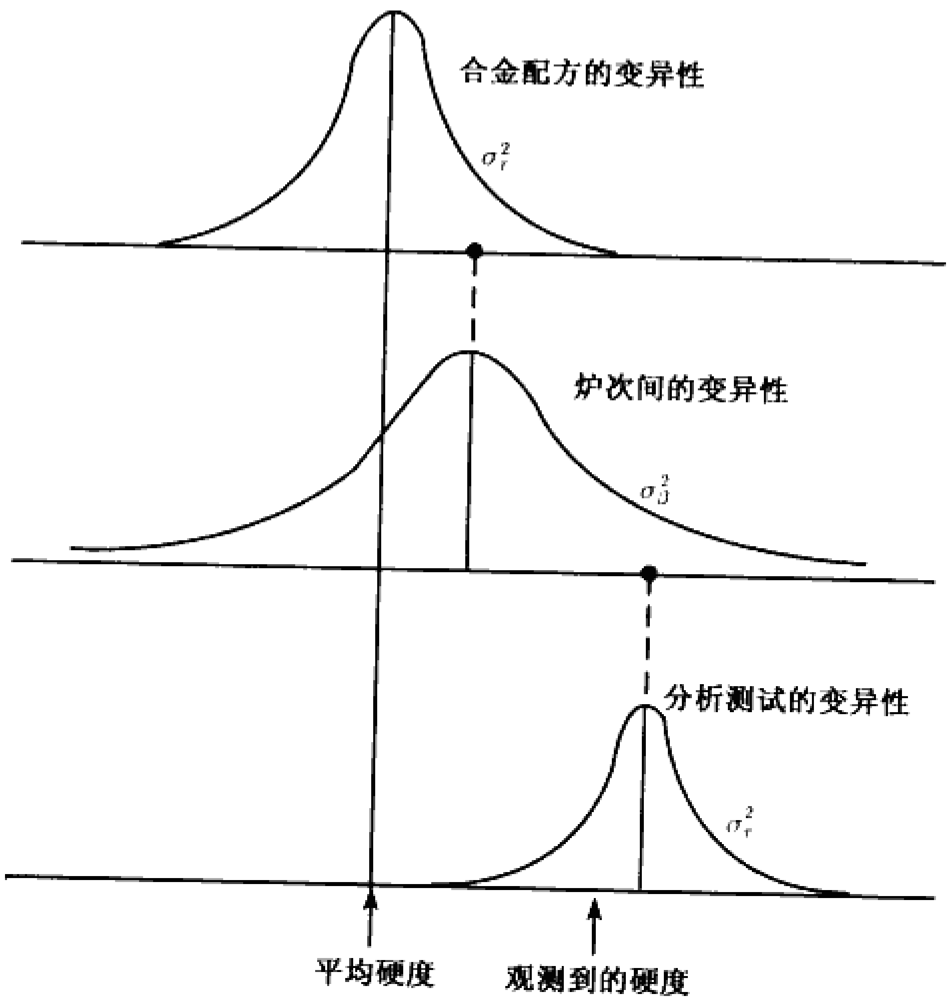


图 14.6　三级嵌套设计例子的方差来源

这个例子说明了嵌套设计是如何在分析过程中用来识别输出中变异性的主要来源的. 例如, 如果合金配方方差分量较大, 就意味着总硬度变异性可以只用某种合金配方来减少.

m 级嵌套设计的平方和的计算法及方差分析和 14.1 节介绍的分析方法相同. 例如, 三级嵌套设计的方差分析见表 14.7, 平方和的计算公式亦列在此表内. 它们是二级嵌套设计计算公式的简单推广.

表 14.7 三级嵌套设计的方差分析

方差来源	平方和	自由度	均方
A	$bcn\sum_i(\bar{y}_{i\ldots}-\bar{y}_{\ldots\ldots})^2$	$a-1$	MS_A
B(A内的)	$cn\sum_i\sum_j(\bar{y}_{ij..}-\bar{y}_{\ldots\ldots})^2$	$a(b-1)$	$MS_{B(A)}$
C(B内的)	$n\sum_i\sum_j\sum_k(\bar{y}_{ijk.}-\bar{y}_{\ldots\ldots})^2$	$ab(c-1)$	$MS_{C(B)}$
误差	$\sum_i\sum_j\sum_k\sum_l(y_{ijkl}-\bar{y}_{ijk.})^2$	$abc(n-1)$	MS_E
总和	$\sum_i\sum_j\sum_k\sum_l(y_{ijkl}-\bar{y}_{\ldots\ldots})^2$	$abcn-1$	

为了确定恰当的检验统计量, 必须用第 13 章的方法来求期望均方. 例如, 如果因子 A 和 B 是固定的, 因子 C 是随机的, 则可以像表 14.8 所列的那样来推导出期望均方. 这张表给出了此种情况的恰当检验统计量.

表 14.8 A 和 B 固定、C 随机的三级嵌套设计的期望均方

模型项	期望均方
τ_i	$\sigma^2+n\sigma_\gamma^2+\dfrac{bcn\sum\tau_i^2}{a-1}$
$\beta_{j(i)}$	$\sigma^2+n\sigma_\gamma^2+\dfrac{cn\sum\sum\beta_{j(i)}^2}{a(b-1)}$
$\gamma_{k(ij)}$	$\sigma^2+n\sigma_\gamma^2$
$\varepsilon_{l(ijk)}$	σ^2

14.3 含被套因子和交叉因子的设计

有时, 在多因子实验中, 一些因子是交叉的, 而另一些因子是被套的. 有时称这类设计为**嵌套析因设计**(nested-factorial design). 下面的例子用于说明有 3 个因子的这种设计的统计分析.

例 14.2 一位工业工程师研究在印刷电路板上手工嵌入电子元件, 以求改进装配速度. 他设计了看来有前景的 3 种装配夹具和 2 种工作场所布局. 需要操作工来进行装配, 人们决定对每种夹具–布局组合随机选用 4 位操作工. 但是, 因为工作场所位于车间内不同的位置, 人们难于对每种布局选用同样的4 位操作工进行工作. 因此, 给布局 1 选用的 4 位操作工和给布局 2 选用的 4 位操作工不相同. 因为只有 3 种夹具和 2 种布局, 而操作工是随机选取的, 这是一个**混合模型**. 此设计的处理组合以随机顺序进行有两次重复的实验. 以秒为单位测得装配时间如表 14.9 所示.

在这一实验中, 操作工被嵌套在布局的水平内, 而夹具和布局是交叉的. 这样一来, 此设计既有被套因子又有交叉因子, 其**线性模型**是

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_{k(j)} + (\tau\beta)_{ij} + (\tau\gamma)_{ik(j)} + \varepsilon_{(ijk)l} \quad \begin{cases} i = 1,2,3 \\ j = 1,2 \\ k = 1,2,3,4 \\ l = 1,2 \end{cases} \tag{14.14}$$

表 14.9　例 14.2 的装配时间数据

操作工	布局 1				布局 2				$y_{i...}$
	1	2	3	4	1	2	3	4	
夹具 1	22	23	28	25	26	27	28	24	404
	24	24	29	23	28	25	25	23	
夹具 2	30	29	30	27	29	30	24	28	447
	27	28	32	25	28	27	23	30	
夹具 3	25	24	27	26	27	26	24	28	401
	21	22	25	23	25	24	27	27	
操作工总和 $y_{.jk.}$	149	150	171	149	163	159	151	160	
布局总和 $y_{.j..}$	619				633				1 252=$y_{....}$

其中 τ_i 是第 i 种夹具的效应, β_j 是第 j 种布局的效应, $\gamma_{k(j)}$ 是在第 j 种布局水平内第 k 个操作工的效应, $(\tau\beta)_{ij}$ 是夹具 × 布局交互作用, $(\tau\gamma)_{ik(j)}$ 是布局内的夹具 × 操作工交互作用, $\varepsilon_{(ijk)l}$ 是通常的误差项. 注意, 不存在布局 × 操作工交互作用, 因为每个操作工不能使用所有的布局. 同理, 不存在三因子交互作用夹具 × 布局 × 操作工. 表 14.10 给出了用第 13 章中的方法并假定是在约束混合模型下得到的期望均方. 查阅这张表就可求得任一效应或交互作用的恰当检验统计量.

表 14.10　例 14.2 的期望均方

模型项	期望均方
τ_i	$\sigma^2 + 2\sigma^2_{\tau\gamma} + 8\sum\tau_i^2$
β_j	$\sigma^2 + 6\sigma^2_{\gamma} + 24\sum\beta_j^2$
$\gamma_{k(j)}$	$\sigma^2 + 6\sigma^2_{\gamma}$
$(\tau\beta)_{ij}$	$\sigma^2 + 2\sigma^2_{\tau\gamma} + 4\sum\sum(\tau\beta)^2_{ij}$
$(\tau\gamma)_{ik(j)}$	$\sigma^2 + 2\sigma^2_{\tau\gamma}$
$\varepsilon_{(ijk)l}$	σ^2

完整的方差分析见表 14.11. 可以看到, 装配夹具是显著的, 布局内的操作工也有显著差异. 布局内夹具和操作工之间也有显著的交互作用, 这表明, 不同夹具的效应对所有操作工说来也不是相同的. 工作场所布局看起来对装配时间只有很小的效应. 因此, 要使装配时间最小, 我们应该集中于夹具类型 1 和 3. (注意表 14.9 的夹具总和, 夹具类型 1 和 3 比类型 2 较小. 夹具类型的均值差可以用多重比较法来正式检验.) 还有, 操作工和夹具之间的交互作用意味着使用相同的夹具时, 某些操作工比另外的操作工效率高. 这类操作工–夹具效应多半可单独处理, 比如对操作工可以重新培训以提高他们的技能.

表 14.11　例 14.2 的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
夹具 (F)	82.80	2	41.40	7.54	0.01
布局 (L)	4.08	1	4.09	0.34	0.58
操作工 (布局内的), $O(L)$	71.91	6	11.99	5.15	<0.01
FL	19.04	2	9.52	1.73	0.22
$FO(L)$	65.84	12	5.49	2.36	0.04
误差	56.00	24	2.33		
总和	299.67	47			

计算

一些统计软件包可以很容易地分析套析因设计, 包括 Minitab 和 SAS. 在约束形式的混合模型假定下, 表 14.12 给出了例 14.2 的 Minitab(Balanced ANOVA) 输出. 位于表 14.12 的下半部分的期望均方与表 14.10 给出的结果一致. $Q[1]$, $Q[3]$, $Q[4]$ 分别是布局、夹具以及布局 × 夹具的固定效应. 方差分量的估计是:

$$\text{操作工 (布局)}: \sigma_\gamma^2 = 1.609$$

$$\text{夹具} \times \text{操作工 (布局)}: \sigma_{\tau\gamma}^2 = 1.576$$

$$\text{误差}: \sigma^2 = 2.333$$

表 14.12 使用约束模型对例 14.2 的 Minitab Balanced ANOVA 的分析

Analysis of Variance (Balanced Designs)

Factor	Type	Levels	Values			
Layout	fixed	2	1	2		
Operator (Layout)	random	4	1	2	3	4
Fixture	fixed	3	1	2	3	

Analysis of Variance for Time

Source	DF	SS	MS	F	P
Layout	1	4.083	4.083	0.34	0.581
Operator (Layout)	6	71.917	11.986	5.14	0.002
Fixture	2	82.792	41.396	7.55	0.008
Layout*Fixture	2	19.042	9.521	1.74	0.218
Fixture*Operator (Layout)	12	65.833	5.486	2.35	0.036
Error	24	56.000	2.333		
Total	47	299.667			

Source	Variance component	Error term	Expected Mean Square for Each Term (using restricted model)
1 Layout		2	(6) + 6(2) + 24Q[1]
2 Operator (Layout)	1.609	6	(6) + 6(2)
3 Fixture		5	(6) + 2(5) + 16Q[3]
4 Layout*Fixture		5	(6) + 2(5) + 8Q[4]
5 Fixture*Operator (Layout)	1.576	6	(6) + 2(5)
6 Error	2.333		(6)

表 14.13 给出了 Minitab 使用**无约束**形式的混合模型对例 14.2 的分析. 在这张表的下半部分给出的期望均方与使用约束模型得到的报告结果略有差异, 于是对操作工 (布局) 因子的检验统计量的构造也略有不同. 特别地, 在约束模型中操作工 (布局) 的 F 比的分母是夹具 × 操作

工 (布局) 的交互作用 (误差项自由度是 12), 而在无约束模型中 F 比的分母是布局 × 夹具的交互作用 (误差项自由度是 2). 因为 $MS_{\text{布局}\times\text{夹具}} > MS_{\text{夹具}\times\text{操作工 (布局)}}$, 它的自由度比较少, 所以布局效应内的操作工只在 12%水平下是显著的 (在约束模型分析中相应的 P 值是 0.002). 并且, 方差分量估计 $\hat{\sigma}_{\gamma}^2 = 1.083$ 较小. 然而, 因为有着很大的夹具效应和显著的夹具 × 操作工 (布局) 交互作用, 我们仍然怀疑这里存在着操作工效应, 所以这个例子得到的实际结论不会因为选择了混合模型的约束形式或者无约束形式而受到很大的影响. $Q[1,4]$ 和 $Q[3,4]$ 是包含布局 × 夹具这个交互效应的固定类型的平方项.

表 14.13 使用无约束模型对例 14.2 的 Minitab Balanced ANOVA 的分析

Analysis of Variance (Balanced Designs)

Factor	Type	Levels	Values			
Layout	fixed	2	1	2		
Operator (Layout)	random	4	1	2	3	4
Fixture	fixed	3	1	2	3	

Analysis of Variance for Time

Source	DF	SS	MS	F	P
Layout	1	4.083	4.083	0.34	0.581
Operator (Layout)	6	71.917	11.986	2.18	0.117
Fixture	2	82.792	41.396	7.55	0.008
Layout*Fixture	2	19.042	9.521	1.74	0.218
Fixture*Operator (Layout)	12	65.833	5.486	2.35	0.036
Error	24	56.000	2.333		
Total	47	299.667			

Source	Variance component	Error term	Expected Mean Square for Each Term (using restricted model)
1 Layout		2	(6) + 2(5) + 6(2) + Q[1,4]
2 Operator (Layout)	1.083	5	(6) + 2(5) + 6(2)
3 Fixture		5	(6) + 2(5) + Q[3,4]
4 Layout*Fixture		5	(6) + 2(5) + Q[4]
5 Fixture*Operator (Layout)	1.576	6	(6) + 2(5)
6 Error	2.333		(6)

如果手头没有像 SAS 或 Minitab 这类软件, 分析析因实验的计算程序也可以用来分析有被套因子和交叉因子的设计. 例如, 例 14.2 可考虑为三因子析因设计, 以夹具 (F)、操作工 (O) 以

及布局 (L) 作为因子. 然后, 将析因设计分析所得的某些平方和与自由度合并起来, 构成被套因子和交叉因子设计所需的合适量, 如下表所示.

析因分析		套析因分析	
平方和	自由度	平方和	自由度
SS_F	2	SS_F	2
SS_L	1	SS_L	1
SS_{FL}	2	SS_{FL}	2
SS_O	3		
SS_{LO}	3	$SS_{O(L)} = SS_O + SS_{LO}$	6
SS_{FO}	6		
SS_{FOL}	6	$SS_{FO(L)} = SS_{FO} + SS_{FOL}$	12
SS_E	24	SS_E	24
SS_T	47	SS_T	47

14.4 裂区设计

在有些多因子析因实验中, 我们也许不能使试验顺序做到完全随机化. 这种情况导致了析因设计的一种推广, 叫做**裂区设计**(split-plot design).

举个例子, 有家造张厂针对 3 种不同的纸浆配制方法和 4 种不同的纸浆蒸煮温度, 想要研究这两个因子对纸张的抗拉强度的效应. 作为析因实验, 每次重复需做 12 个观测值, 而且实验者已经决定要做 3 次重复. 但是, 实验室只能容许每天做 12 个试验, 所以, 实验者决定分为 3 天来做, 每天做一次, 并且把天或重复看作为区组. 任何一天, 他进行如下的实验. 用所研究的 3 种方法之一生产出来一批纸浆. 然后, 将这批纸浆分为 4 个样本, 每个样本用 4 种温度之一来蒸煮. 然后, 第 2 批纸浆用 3 种方法的另一种制造出来. 第 2 批纸浆也分为 4 个样本, 也用 4 种温度来试验. 再用第 3 种方法生产一批纸浆, 重复这一过程. 所得数据如表 14.14 所示.

表 14.14 纸张抗拉强度实验

纸浆配制方法	重复 (或区组)1			重复 (或区组)2			重复 (或区组)3		
	1	2	3	1	2	3	1	2	3
温度 (°F)									
200	30	34	29	28	31	31	31	35	32
225	35	41	26	32	36	30	37	40	34
250	37	38	33	40	42	32	41	39	39
275	36	42	36	41	40	40	40	44	45

起初, 可以把这看作是配制方法 (因子 A) 有 3 个水平、温度 (因子 B) 有 4 个水平的随机化区组析因实验. 如果是这种情况的话, 则区组内的实验顺序应该是完全随机化的. 也就是说, 在一个区组内, 我们应该随机地选择一种处理组合 (一种配制方法和一种温度) 并得到一个观测值, 然后又随机地选取另一处理组合得到第 2 个观测值, 等等, 直到求得这个区组内的 12 个观测值为止. 但是, 现在实验者并不按这种方式收集数据. 他做好一批纸浆, 并从这批纸浆中对所

有 4 种温度求得观测值. 因为这样配制纸浆又省工又省料, 是收集数据的最适当可行的方法. 一个完全随机化析因实验需要 36 批纸浆, 这是完全不现实的. 裂区设计对每个区组 (重复) 只需要 3 批纸浆, 从而在这个例子中总共需要 9 批. 显然裂区设计能在很大程度上提高实验效率.

我们的这个例子使用的是裂区设计. 裂区设计中的每个重复或区组分为 3 个部分, 叫做**全区**, 配制方法叫做**全区处理**或**主处理**. 每个全区分为 4 个部分, 叫做**子区**(或**裂区**), 一种温度分配一个子区. 温度叫做**子区处理**. 如果有不可控因子或未设计的因子出现, 并且这些不可控的因子随着纸浆配制方法的改变而变化时, 则未设计因子对响应的任何效应将完全与纸浆配制方法的效应相混. 因为在裂区设计中, 全区处理与全区相混而子区处理是不相混的, 所以, 只要有可能, 最好是把我们最感兴趣的因子安排在子区中.

这是将裂区设计用于工业背景中的一个典型例子, 注意到两个因子用在不同的时间上, 所以一个裂区设计可以看作是两个实验的 "合并" 或相互添加. 一个 "实验" 把全区因子应用在大的实验单元上 (或这些因子的水平很难改变), 另一 "实验" 把子区因子应用在较小的实验单元上 (即这个因子的水平改变比较容易).

裂区设计的**线性模型**是

$$y_{ijk} = \mu+\tau_i+\beta_j+(\tau\beta)_{ij}+\gamma_k+(\tau\gamma)_{ik}+(\beta\gamma)_{jk}+(\tau\beta\gamma)_{ijk}+\varepsilon_{ijk} \quad \begin{cases} i = 1,\ 2,\ \cdots,\ r \\ j = 1,\ 2,\ \cdots,\ a \\ k = 1,\ 2,\ \cdots,\ b \end{cases} \tag{14.15}$$

其中 τ_i, β_j, $(\tau\beta)_{ij}$ 代表全区并分别对应于区组 (或重复)、主效应 (因子 A) 和**全区误差**[重复 (或区组)$\times A$); 而 γ_k, $(\tau\gamma)_{ik}$, $(\beta\gamma)_{jk}$, $(\tau\beta\gamma)_{ijk}$ 代表子区并分别对应于子区处理 (因子 B)、重复 (或区组)$\times B$、AB 交互作用以及**子区误差**(区组 $\times AB$). 注意, 全区误差是重复 (或区组) 交互作用, 子区误差是三因子交互作用区组 $\times AB$. 这些因子的平方和可以作为单次重复的三向方差分析来计算.

当重复 (即区组) 随机、主处理和子区处理固定时, 裂区设计的期望均方见表 14.15. 全区中的主因子 (A) 是针对全区误差来检验的, 而子处理 (B) 是针对重复 (或区组)$\times$ 子处理交互作用来检验的. AB 交互作用是针对子区误差来检验的. 对重复 (或区组) 效应 (A) 或者重复 (或区组)$\times$ 子处理 (AC) 交互作用没有检验法.

表 14.14 的抗拉强度数据的方差分析概括在表 14.16 中. 因为配制方法和温度是固定的而重复是随机的, 所以应用表 14.15 的期望均方. 配制方法的均方与全区误差的均方相比较, 温度的均方与重复 (或区组)$\times$ 温度 (AC) 的均方相比较. 最后, 配制方法 $\times$ 温度的均方针对子区误差来检验. 配制方法和温度对抗拉强度都有显著的效应, 并且它们的交互作用是显著的.

从表 14.16 中看到, 子区误差 (4.24) 小于全区误差 (9.07). 这是裂区设计的普遍情形, 因为子区比起全区来说一般更为齐性. 这导致了实验的**两种不同的误差结构**. 由于子区处理以更大的精确度进行比较, 所以, 只要有可能, 应该把我们最感兴趣的处理安排在子区中.

有些作者对裂区设计提出了一个略有不同的统计模型, 比如

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_k + (\beta\gamma)_{jk} + \varepsilon_{ijk} \quad \begin{cases} i = 1,\ 2,\ \cdots,\ r \\ j = 1,\ 2,\ \cdots,\ a \\ k = 1,\ 2,\ \cdots,\ b \end{cases} \tag{14.16}$$

表 14.15 裂区设计的期望均方

	模型项	期望均方
全区	τ_i	$\sigma^2 + ab\sigma_\tau^2$
	β_j	$\sigma^2 + b\sigma_{\tau\beta}^2 + \frac{rb\sum\beta_j^2}{a-1}$
	$(\tau\beta)_{ij}$	$\sigma^2 + b\sigma_{\tau\beta}^2$
子区	γ_k	$\sigma^2 + a\sigma_{\tau\gamma}^2 + \frac{ra\sum\gamma_k^2}{(b-1)}$
	$(\tau\gamma)_{ik}$	$\sigma^2 + a\sigma_{\tau\gamma}^2$
	$(\beta\gamma)_{jk}$	$\sigma^2 + \sigma_{\tau\beta\gamma}^2 + \frac{r\sum\sum(\beta\gamma)_{jk}^2}{(a-1)(b-1)}$
	$(\tau\beta\gamma)_{ijk}$	$\sigma^2 + \sigma_{\tau\beta\gamma}^2$
	$\varepsilon_{(ijk)h}$	σ^2(不可估计的)

表 14.16 采用表 14.14 的抗拉强度数据的裂区设计的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
重复 (或区组)	77.55	2	38.78		
配置方法 (A)	128.39	2	64.20	7.08	0.05
全区误差 [重复 (或区组)×A]	36.28	4	9.07		
温度 (B)	434.08	3	144.69	41.94	< 0.01
重复 (或区组)×B	20.67	6	3.45		
AB	75.17	6	12.53	2.96	0.05
子区误差 [重复 (或区组)×AB]	50.83	12	4.24		
总和	822.97	35			

在这个模型中 $(\tau\beta)_{ij}$ 仍然是全区误差, 但是区组 $\times B$ 和区组 $\times AB$ 的交互作用实质上是和 ε_{ijk} 合并形成了子区误差. 我们记子区误差项 ε_{ijk} 的方差为 σ_ε^2, 并作与模型 [(14.15) 式] 相同的假设, 那么期望均方变成

因　子	$E(MS)$
τ_i(重复或区组)	$\sigma_\varepsilon^2 + ab\sigma_\tau^2$
$\beta_j(A)$	$\sigma_\varepsilon^2 + b\sigma_{\tau\beta}^2 + \frac{rb\sum\beta_j^2}{a-1}$
$(\tau\beta)_{ij}$	$\sigma_\varepsilon^2 + b\sigma_{\tau\beta}^2$(全区误差)
$\gamma_k(B)$	$\sigma_\varepsilon^2 + \frac{ra\sum\gamma_k^2}{ab-1}$
$(\beta\gamma)_{jk}(AB)$	$\sigma_\varepsilon^2 + \frac{r\sum\sum(\beta\gamma)_{jk}^2}{(a-1)(b-1)}$
ε_{ijk}	σ_ε^2(子区误差)

注意, 现在子区处理 (B) 和 AB 交互作用都是针对子区误差期望均方进行检验的. 如果人们觉得重复 (或区组)×B 交互作用和重复 (或区组)×AB 交互作用都可以忽略这一假设相当合理, 那么这个模型是完全令人满意的.

表 14.17 给出了表 14.14 裂区设计的基于 (14.16) 式的模型的 Minitab 输出. 用 Balanced ANOVA 模块计算时采用了约束模型. (重复或区组是一个随机因子, 而纸浆配制方法和温度是固定因子, 所以这是一个混合模型分析.) 这个分析得到的结果与表 14.16 中原来的 ANOVA 分析结果非常接近. 注意, 因子 "重复" 是针对子区误差进行检验的.

表 14.17 表 14.14 裂区设计的 ANOVA 分析 (另一个模型)

ANOVA: 强度与重复、配制方法、温度

Factor	Type	Levels	Values
Replicate	random	3	1, 2, 3
Prep Meth	fixed	3	1, 2, 3
Temp	fixed	4	200, 225, 250, 275

Analysis of Variance for Strenght

Source	DF	SS	MS	F	P
Replicate	2	77.556	38.778	9.76	0.001
Prep Meth	2	128.389	64.194	7.08	0.049
Replicate*Preo Meth	4	36.278	9.069	2.28	0.100
Temp	3	434.083	144.694	36.43	0.000
Prep Meth*Temp	6	75.167	12.528	3.15	0.027
Error	18	71.500	3.972		
Total	35	822.972			

S = 1.99304 R-Sq = 91.31% R-Sq (adj) = 83.11%

Source	Variance component	Error term	Expected Mean Square for Each Term (using restricted model)
1 Replicate	2.900	6	(6) + 12 (1)
2 Prep Meth		3	(6) + 4 (3) + 12 Q[2]
3 Replicate*Prep Meth	1.274	6	(6) + 4 (3)
4 Temp		6	(6) + 9 Q[4]
5 Prep Meth*Temp		6	(6) + 3 Q[5]
6 Error	3.972		(6)

裂区设计有农业的传统, 全区通常是大片土地, 子区是在大片土地中的小片土地. 例如, 一农作物的几种不同的品种可以种在不同的土地上 (全区), 一种品种种一块土地. 然后, 每块土地可分为比方说 4 个子区, 每个子区施用不同类型的肥料. 这里, 农作物品种是主处理, 不同的肥料是子处理.

不管它的农业基础, 裂区设计在很多科学实验和工业实验中也都有用. 在这类实验的策划过程中, 通常去找出一些需用较大实验单元的因子, 而另一些因子需用较小的实验单元, 如同上面描述的抗拉强度问题那样. 另外, 我们有时发现, 完全随机化是不适宜的, 因为, 有些因子比起其他因子来说更难于改变它的水平. 难于改变的因子形成全区, 而易于改变的因子在子区中做试验.

从原则上说来, 我们必须小心地考虑如何进行实验, 并将关于随机化的所有约束结合到分析方法中去. 我们用修正过的第 6 章的眼睛聚焦时间的例子来说明这一点. 假定只有两个因子: 视觉敏锐度 (A) 和照度 (B). 视觉敏锐度有 a 个水平, 照度有 b 个水平, 进行 n 次重复的析因实验, 需要以随机顺序取得 abn 个观测值. 但是, 相当难于调节这两个因子至不同的水平上, 所以, 实验者决定通过将仪器调节至 a 个敏锐度之一和 b 个照度之一并立刻做 n 次重复试验以取得 n 个观测值. 在析因设计中, 误差实际上代表系统的干扰或噪音加上被实验者再现同样的聚焦时间的能力. 析因设计的模型可写为

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \phi_{ijk} + \theta_{ijk} \quad \begin{cases} i = 1,\ 2,\ \cdots,\ a \\ j = 1,\ 2,\ \cdots,\ b \\ k = 1,\ 2,\ \cdots,\ n \end{cases} \tag{14.17}$$

其中 ϕ_{ijk} 代表系统的干扰或噪音, 这是由 "实验误差" 产生的 (也就是说, 我们在不同的试验、环境条件变异性及类似的原因中, 实际上不能做到完完全全地重复敏锐度和照度的同一水平), θ_{ijk} 代表被试验者的 "再现性误差". 通常, 我们把这些分量组合为一个总的误差项, 记作 $\varepsilon_{ijk} = \phi_{ijk} + \theta_{ijk}$. 假定 $\mathrm{V}(\varepsilon_{ijk}) = \sigma^2 = \sigma_\phi^2 + \sigma_\theta^2$. 现在, 在此析因设计中, 误差均方的期望为 $\sigma^2 = \sigma_\phi^2 + \sigma_\theta^2$, 有 $ab(n-1)$ 个自由度.

如果像前面第 2 个设计那样限制随机化, 则方差分析中的 "误差" 均方是 "再现性误差"σ_θ^2 的估计量, 有 $ab(n-1)$ 个自由度, 但得不到关于 "实验误差"σ_ϕ^2 的信息. 这样一来, 在第 2 个设计中的误差的均方是太小了, 因此, 我们经常会错误地否定零假设. 正如 John(1971) 所指出的, 这一设计类似于有 ab 个全区, 每个全区分为 n 个子区, 没有子处理的裂区设计. 这种情况也类似于 Ostle(1963) 所描述的**子抽样**. 假定 A 和 B 是固定的, 我们得到此时的期望均方是

$$\begin{aligned} &\mathrm{E}(MS_A) = \sigma_\theta^2 + n\sigma_\phi^2 + \frac{bn\sum\tau_i^2}{a-1}, \qquad \mathrm{E}(MS_{AB}) = \sigma_\theta^2 + n\sigma_\phi^2 + \frac{n\sum\sum(\tau\beta)_{ij}^2}{(a-1)(b-1)} \\ &\mathrm{E}(MS_B) = \sigma_\theta^2 + n\sigma_\phi^2 + \frac{an\sum\beta_j^2}{b-1}, \qquad \mathrm{E}(MS_E) = \sigma_\theta^2 \end{aligned} \tag{14.18}$$

因此, 除非交互作用可忽略, 不然就不能检验主效应. 这种情况实际上是每单元有一个观测值的双向方差分析. 如果两个因子都是随机的, 则主效应可以针对 AB 交互作用来检验. 如果只有一个因子是随机的, 则固定因子可以针对 AB 交互作用来检验.

一般来说, 如果谁在分析析因设计时, 得出的结论是所有的主效应和交互作用都是显著的, 那他就应该小心地检查一下**实验实际上是如何进行的**. 有可能是在分析时没有考虑模型的随机化约束, 因而不应该把数据作为析因设计来分析.

14.5 裂区设计的其他变形

14.5.1 多于两个因子的裂区设计

有时我们会遇到全区或子区含有两个或更多的因子，它们排成了析因的结构. 作为一个例子，考虑在炉子上给硅晶片生成氧化层的实验，所关心的响应变量是氧化层的厚度和均匀度. 有 4 个设计因子：温度 (A)，气流 (B)，时间 (C)，晶片在炉子里的位置 (D). 实验者计划用一个 2^4 析因设计，重复两次 (共 32 次试验). 现在因子 A 与 B(温度和气流) 很难改变，而 C 和 D(时间和晶片位置) 很容易改变. 这样就产生了一个如图 14.7 所示的裂区设计. 注意实验的两次重复都被分裂为 4 个全区，每个全区都包含温度和气流的一种设置组合. 当 A 和 B 的水平被选定后，每个全区又分为 4 个子区，并对时间和晶片位置做了一个 2^2 析因实验，其中子区中的处理组合的试验顺序是随机的. 每次重复中温度和气流只改变了 4 次，而时间和晶片位置的水平是完全随机的.

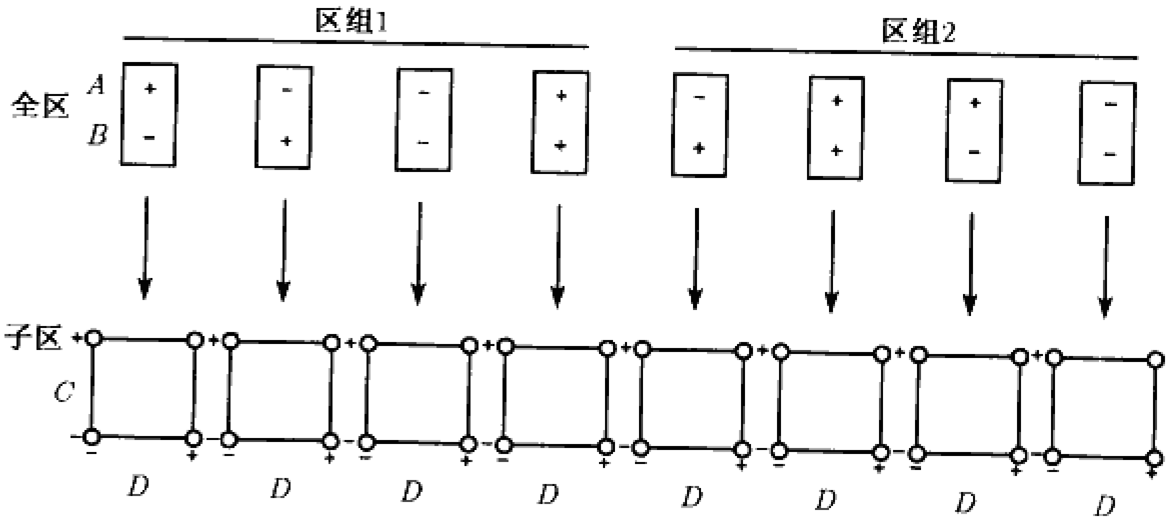


图 14.7 一个四因子裂区设计，两个因子在全区，两个因子在子区

与 (14.16) 式一致，此实验的一个模型是

$$
\begin{aligned}
y_{ijklm} = {} & \mu + \tau_i + \beta_j + \gamma_k + (\beta\gamma)_{jk} + \theta_{ijk} + \delta_l + \lambda_m + (\delta\lambda)_{lm} + (\beta\delta)_{jl} + (\beta\lambda)_{jm} \\
& + (\gamma\delta)_{kl} + (\delta\lambda)_{lm} + (\beta\gamma\delta)_{jkl} + (\beta\gamma\lambda)_{jkm} + (\beta\delta\lambda)_{jlm} + (\gamma\delta\lambda)_{klm} \\
& + (\beta\gamma\delta\lambda)_{jklm} + \varepsilon_{ijklm} \qquad \begin{cases} i = 1,2 \\ j = 1,2 \\ k = 1,2 \\ l = 1,2 \\ m = 1,2 \end{cases}
\end{aligned} \tag{14.19}
$$

其中 τ_i 表示重复效应，β_j 和 γ_k 表示全区主效应，θ_{ijk} 是全区误差，δ_l 和 λ_m 表示子区主效应，ε_{ijklm} 是子区误差. 这里包含了 4 个设计因子的所有交互作用. 表 14.18 给出了在重复是随机的、所有设计因子是固定的假设下，这个设计的方差分析. 该表中 σ_θ^2 和 σ_ε^2 分别表示全区和子区的误差方差，σ_τ^2 是区组效应的方差，(为简单起见) 我们用大写拉丁字母标记固定效应. 将全区主效应和交互作用与全区误差比较进行检验，将子区效应和所有其他交互作用与子区误差比较

进行检验. 如果某些设计因子是随机的, 则检验统计量就会不一样. 有时没有精确的 F 检验, 而要使用 (第 13 章中介绍的)Satterthwaite 方法.

表 14.18 裂区设计的简要分析, 其中 A 和 B 是全区因子, C 和 D 是子区因子 (参见图 14.7)

方差来源	平方和	自由度	期望均方	方差来源	平方和	自由度	期望均方
重复 (τ_i)	$SS_{重复}$	1	$\sigma_\varepsilon^2+16\sigma_\tau^2$	AD	SS_{AD}	1	σ_ε^2+AD
$A(\beta_k)$	SS_A	1	$\sigma_\varepsilon^2+8\sigma_\theta^2+A$	BD	SS_{BD}	1	σ_ε^2+BD
$B(\gamma_k)$	SS_B	1	$\sigma_\varepsilon^2+8\sigma_\theta^2+B$	ABC	SS_{ABC}	1	σ_ε^2+ABC
AB	SS_{AB}	1	$\sigma_\varepsilon^2+8\sigma_\theta^2+AB$	ABD	SS_{ABD}	1	σ_ε^2+ABD
全区误差 (θ_{ijk})	SS_{WP}	3	$\sigma_\varepsilon^2+8\sigma_\theta^2$	ACD	SS_{ACD}	1	σ_ε^2+ACD
$C(\delta_l)$	SS_C	1	σ_ε^2+C	BCD	SS_{BCD}	1	σ_ε^2+BCD
$D(\lambda_m)$	SS_D	1	σ_ε^2+D	$ABCD$	SS_{ABCD}	1	$\sigma_\varepsilon^2+ABCD$
CD	SS_{CD}	1	σ_ε^2+CD	子区误差 (ε_{ijklm})	SS_{SP}	12	σ_ε^2
AC	SS_{AC}	1	σ_ε^2+AC	总和	SS_T	31	
BC	SS_{BC}	1	σ_ε^2+BC				

在裂区结构中, 含有 3 个或 3 个以上因子的析因实验将会是一个相当大的实验. 另一方面, 裂区结构通常使这样的一个大型实验更容易实施: 例如在氧化炉例子中, 实验者只需要将难以改变的因子 (A 和 B) 改变 8 次, 所以一个 32 次的实验也并非不合理.

当实验中因子的个数增多时, 实验者可以考虑采用以裂区方式安排的分式析因实验. 作为一个例子, 考虑思考题 8.7 中所提到的实验. 这是一个 2^{5-1} 分式析因实验, 用来研究热处理过程变量对汽车簧片高度的影响, 因子 $A=$ 传递时间, $B=$ 加热时间, $C=$ 淬火油温, $D=$ 炉温, $E=$ 保持时间, 假设因子 A, B, C 很难改变, 而另两个因子 D 和 E 易于改变. 例如, 在生产簧片时, 首先必须通过调节变化因子 A, B, C, 然后在后续的实验中, 将它们固定, 再让因子 D 和 E 变动.

考虑这个实验的一个修正, 这是因为原来的实验者没有采用裂区方法进行这个实验, 而且他们没有用我们下面将要使用的设计生成元. 全区因子用 A,B,C 表示, 子区因子用 D,E 表示. 选择设计生成元: $E=ABCD$. 设计有 8 个全区 (对应着 3 个因子 2 种水平的 8 种组合). 每个全区分为 2 个子区, 每个子区中对因子 D 和 E 的一种组合进行试验. (精确的处理组合通过生成元依赖于全区因子处理组合的符号.)

假设所有三因子交互作用、四因子交互作用、五因子交互作用都可以忽略. 如果这个假设是合理的, 那么可以估计全区中所有全区因子 A,B,C 和它们的交互作用. 如果这个设计有重复, 则可以将这些效应与全区误差比较进行检验. 如果设计是无重复的, 则这些效应可以用正态概率图 (或 Lenth 的方法) 来评估. 也可以估计子区因子 D 和 E 以及它们的交互作用 DE. 全区因子与子区因子的 6 个交互作用 AD,AE,BD,BE,CD,CE 也能估计. 总之, 所有裂区主效应, 或与全区主效应别名的交互作用, 或只含有全区因子的交互作用将与全区误差相比较. 而且, 裂区主效应, 或包含至少一个不与全区主效应别名的裂区因子的交互作用, 或仅包含全区因子的交互作用将与裂区误差相比较. 见 Bisgaard (1992) 的全面讨论. 因此, 在我们的问题中, 所有效应 D,E,DE,AD,AE,BD,BE,CD,CE 都与裂区误差进行了比较. 另外, 它们可以用正态概率图评估.

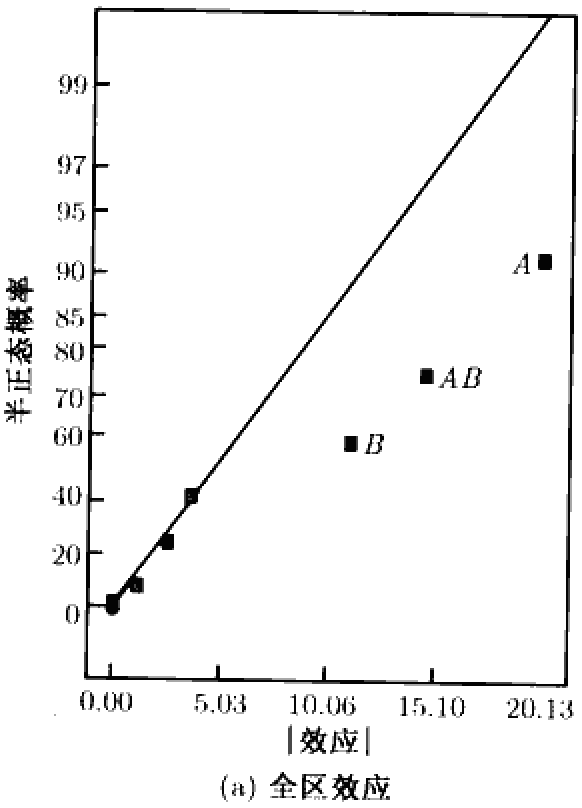
最近, 人们已发表了多篇关于裂区的分式析因问题的论文. 这里着重推荐 Bisgaard(2000). 另外也可以参见 Bingham and Sitter(1999) 以及 Huang, Chen, and Voelkel(1999).

例 14.3 研究单晶片等离子蚀刻工艺中影响均匀性的因子. 相对而言, 蚀刻工具的 3 个因子 (A= 电极间隙, B= 气流, C= 压强) 很难在试验间进行调整. 而另外两个因子 [D= 时间, E = RF(射频) 强度] 很容易在试验间进行调整. 由于可用于测试的晶片数量有限, 实验者打算用一个分式析因实验来研究这 5 个因子. 那些难以改变的因子表明必须考虑裂区设计. 实验者决定用上文所讨论的方法, 即 2^{5-1} 设计, 其中因子 A, B, C 安排在全区, 而 D, E 安排在子区. 设计生成元是 $E = ABCD$. 这样就得到一个含有 8 个全区的 16 次的分式析因设计. 每个全区包含 3 个因子 A, B, C 的完全 2^3 析因设计的 8 个处理组合之一. 每个全区分为两个子区, 每个子区中含有因子 D 和 E 的一种处理组合. 这个设计和所得的均匀度数据见表 14.19. 8 个全区按随机顺序进行试验. 而一旦蚀刻工具上关于因子 A, B, C 的全区安排设置好后, 就把子区中的两个实验做完 (也按随机顺序).

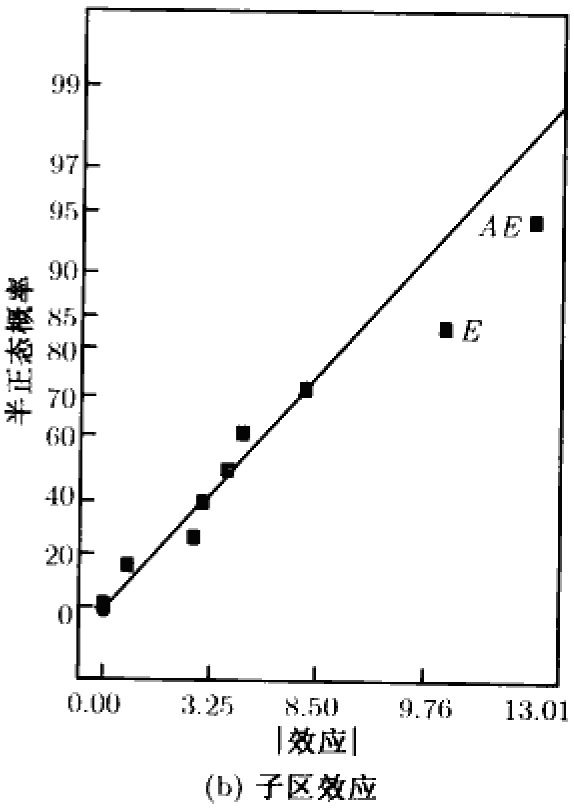
表 14.19 等离子蚀刻工具的 2^{5-1} 裂区实验

全区	全区因子			子区因子		均匀度	全区	全区因子			子区因子		均匀度
	A	B	C	D	E			A	B	C	D	E	
1	−	−	−	−	+	40.85	5	−	−	+	−	−	40.32
	−	−	−	+	−	41.07		−	−	+	+	+	43.34
2	+	−	−	−	−	35.67	6	+	−	+	−	+	62.46
	+	−	−	+	+	51.15		+	−	+	+	−	38.08
3	−	+	−	−	−	41.80	7	−	+	+	−	+	31.99
	−	+	−	+	+	37.01		−	+	+	+	−	41.03
4	+	+	−	−	+	91.09	8	+	+	+	−	−	70.31
	+	+	−	+	−	48.67		+	+	+	+	+	81.03

这个实验的统计分析要把全区和子区因子分开来. 假设所有超过二阶的交互作用都可以忽略. 图 14.8a 是忽略子区因子只看全区因子的效应估计半正态概率图. 从中可以看到因子 A, B 和交互作用 AB 有大的效应. 图 14.8b 是子区效应 D, E 和交互作用 $DE, AD, AE, BD, BE, CD, CE$ 的半正态概率图. 从中看到只有主效应 E 和交互作用 AE 较大.



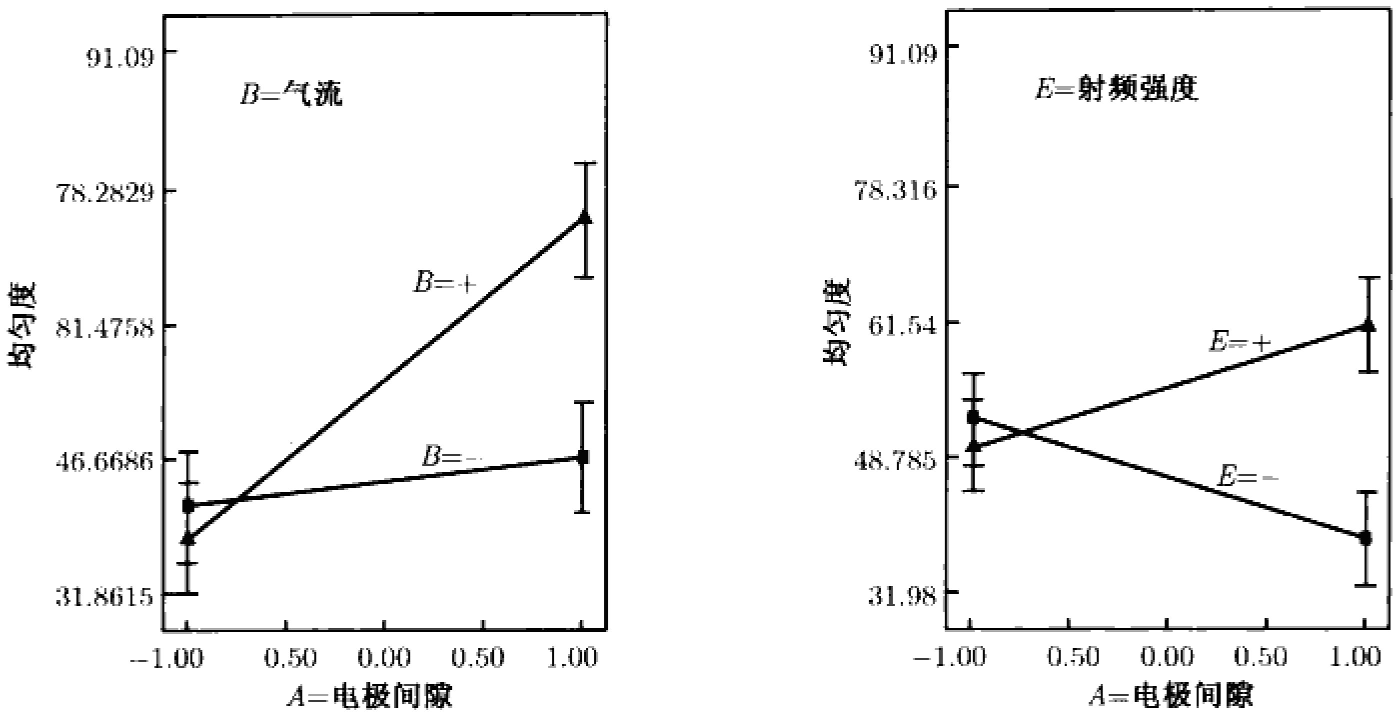
(a) 全区效应



(b) 子区效应

图 14.8 例 14.3 的 2^{5-1} 裂区实验效应的半正态图

图 14.9a 和图 14.9b 分别是 AB 和 AE 的二因子交互作用图. 实验者的目的是最小化均匀度响应, 从图 14.9a 中可以看出, 只要因子 $B=$ 气流处在低水平, 那么因子 $A=$ 电极间隙无论在哪个水平时效果都很好, 如果 B 处在高水平, 那么 A 必须在低水平时才能得到小的均匀度. 图 14.9b 可以看到, 把 $E=$ 射频强度控制在低水平能有效地减小均匀度, 特别是当 A 处在高水平时. 然而若 E 在高水平, 则 A 必须选择低水平. 因此, 这个筛选实验的结果是原来 5 个因子中的 3 个因子对蚀刻均匀度有显著影响. 进一步, 处理组合 A 高、B 低、E 低或者 A 低、B 高、E 高会得到低水平的均匀度响应.



(a) AB 交互作用　　　　(b) AE 交互作用

图 14.9 例 14.3 的 2^{5-1} 裂区实验的二因子交互作用

14.5.2 裂裂区设计

裂区设计的概念可以推广到在一个实验内的随机化约束以任意多个水平出现的情况. 如果随机化约束有两个水平, 则这种方案叫做**裂裂区设计** (split-split-plot design). 下面的例子说明了这种设计.

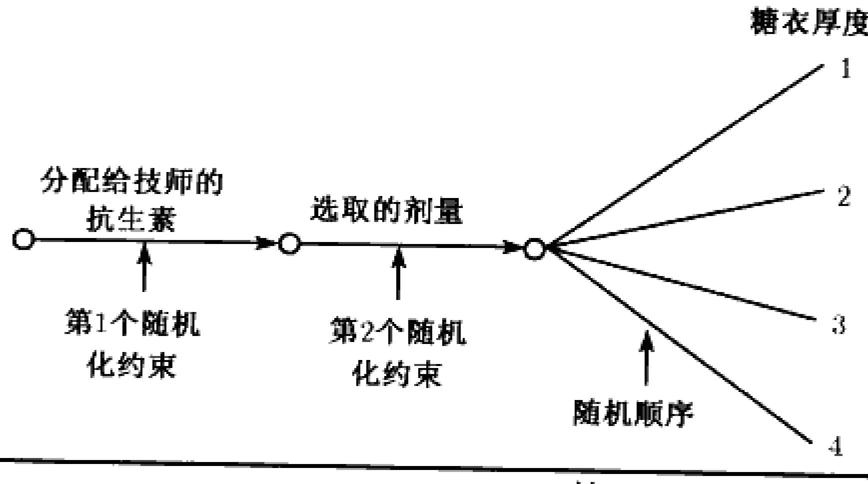
例 14.4 一位研究员研究一种特定类型的抗生素胶囊的吸收时间. 研究员感兴趣的有: 3 位技师, 3 种剂量, 4 种胶囊糖衣厚度. 析因实验设计的每次重复需有 36 个观测值. 实验者决定做 4 次重复, 并且只能每天做一次重复. 这样一来, 一天可以看作一个区组, 在每个重复 (或区组)(天) 内, 进行实验时, 分配一个单位的抗生素给一位技师, 由这位技师来实施 3 种剂量和 4 种糖衣厚度的实验. 一旦配好一种剂量, 就立即在这种剂量下对所有 4 种糖衣厚度进行试验. 然后配制另一种剂量再进行所有 4 种糖衣厚度的试验. 最后, 配制第 3 种剂量并试验 4 种糖衣厚度. 与此同时, 在另外两个实验室, 技师也是用一个单位的抗生素以同样的方法做试验.

我们看到, 在每个重复 (或区组) 内有两个随机化约束: 技师和剂量. 全区对应于技师. 一个单位的抗生素分配给技师的顺序是随机确定的. 剂量构成 3 个子区. 剂量可以随机地分配给子区. 最后, 在一种特定的剂量内, 4 种胶囊糖衣厚度以随机顺序进行试验, 构成 4 个子子区. 糖衣厚度通常叫做子子处理. 因为在实验中有两个随机化约束 (有些作者叫做设计的两次分 "裂"), 所以把这种设计叫做裂裂区设计. 图 14.10 说明了这种设计的随机化约束和实验的布局.

裂裂区设计的线性统计模型是

$$\begin{aligned} y_{ijkh} = &\mu+\tau_i+\beta_j+(\tau\beta)_{ij}+\gamma_k+(\tau\gamma)_{ik}+(\beta\gamma)_{jk}+(\tau\beta\gamma)_{ijk}+\delta_h+(\tau\delta)_{ih}+(\beta\delta)_{jh} \\ &+(\tau\beta\delta)_{ijh}+(\gamma\delta)_{kh}+(\tau\gamma\delta)_{ikh}+(\beta\gamma\delta)_{jkh} \end{aligned}$$

$$+(\tau\beta\gamma\delta)_{ijkh}+\epsilon_{ijkh}\quad\begin{cases}i=1,\ 2,\ \cdots,r\\j=1,\ 2,\ \cdots,a\\k=1,\ 2,\ \cdots,b\\h=1,\ 2,\ \cdots,c\end{cases}\tag{14.20}$$



区组	剂量	技师 1: 1	2	3	技师 2: 1	2	3	技师 3: 1	2	3
1	糖衣厚度	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
2	糖衣厚度	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
3	糖衣厚度	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
4	糖衣厚度	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

图 14.10 一个裂裂区设计

其中 τ_i, β_j, $(\tau\beta)_{ij}$ 代表全区, 分别对应于重复 (或区组)、主处理 (因子 A) 和全区误差 [重复 (或区组) $\times A$]; γ_k, $(\tau\gamma)_{ik}$, $(\beta\gamma)_{jk}$, $(\tau\beta\gamma)_{ijk}$ 代表子区, 分别对应于子区处理 (因子 B), 重复 (或区组)$\times B$, AB 交互作用, 以及子区误差; δ_h 和其余的参数对应于子子区并分别代表子子区处理 (因子 C) 和其余的交互作用. 四因子交互作用 $(\tau\beta\gamma\delta)_{ijkh}$ 叫做子子区误差.

假定重复 (或区组) 是随机的, 其他设计因子是固定的, 表 14.20 给出了此时的期望均方. 审阅一下这张表, 关于主处理、子处理、子子处理以及它们的交互作用的检验是明显的. 注意, 关于区组以及区组的交互作用的检验不存在.

裂裂区设计的统计分析和单次重复的四因子析因设计的分析相像. 各个检验的自由度用通常的方法确定. 为了说明起见, 在例 14.4 中, 我们有 4 个区组, 3 位技师, 3 种剂量, 4 种糖衣厚度, 我们只有 $(r-1)(a-1)=(4-1)(3-1)=6$ 个主区误差自由度用于检验技师, 相对说来,

这是较小的自由度, 实验者可以考虑用附加的重复来提高检验的精确度. 如果有 a 个重复, 则全区误差可以有 $2(r-1)$ 个自由度. 这样, 5 个重复产生 $2(5-1)=8$ 个误差自由度, 6 个重复产生 $2(6-1)=10$ 个误差自由度, 7 个重复产生 $2(7-1)=12$ 个误差自由度, 等等. 因此, 我们也许不想做少于 4 个重复的试验, 因为那只有 4 个误差自由度. 每个附加的重复将增加 2 个误差自由度. 如果能承担 5 个重复的试验的话, 我们就给检验的精确度提高了 1/3 (自由度从 6 增至 8). 而从 5 个重复增至 6 个重复, 精确度将提高 25%. 倘若资源允许, 实验者应该做 5 个或 6 个重复的试验.

表 14.20 裂裂区设计的期望均方

	模型项	期望均方
全区	τ_i	$\sigma^2+abcd\sigma_\tau^2$
	β_j	$\sigma^2+\sigma_{\tau\beta}^2+\dfrac{rbc\sum\beta_j^2}{(a-1)}$
	$(\tau\beta)_{ij}$	$\sigma^2+bc\sigma_{\tau\beta}^2$
子区	γ_k	$\sigma^2+ac\sigma_{\tau\gamma}^2+\dfrac{rac\sum\gamma_k^2}{(b-1)}$
	$(\tau\gamma)_{ik}$	$\sigma^2+ac\sigma_{\tau\gamma}^2$
	$(\beta\gamma)_{jk}$	$\sigma^2+c\sigma_{\tau\beta\gamma}^2+\dfrac{rc\sum\sum(\beta\gamma)_{jh}^2}{(a-1)(b-1)}$
	$(\tau\beta\gamma)_{ijk}$	$\sigma^2+c\sigma_{\tau\beta\gamma}^2$
子子区	δ_h	$\sigma^2+ab\sigma_{\tau\delta}^2+\dfrac{rab\sum\delta_k^2}{(c-1)}$
	$(\tau\delta)_{ih}$	$\sigma^2+ab\sigma_{\tau\delta}^2$
	$(\beta\delta)_{jh}$	$\sigma^2+b\sigma_{\tau\beta\delta}^2+\dfrac{rb\sum\sum(\beta\delta)_{jh}^2}{(a-1)(c-1)}$
	$(\tau\beta\delta)_{ijh}$	$\sigma^2+b\sigma_{\tau\beta\delta}^2$
	$(\gamma\delta)_{kh}$	$\sigma^2+a\sigma_{\tau\gamma\delta}^2+\dfrac{ra\sum\sum(\gamma\delta)_{kh}^2}{(b-1)(c-1)}$
	$(\tau\gamma\delta)_{ikh}$	$\sigma^2+a\sigma_{\tau\gamma\delta}^2$
	$(\beta\gamma\delta)_{jkh}$	$\sigma^2+\sigma_{\tau\beta\gamma\delta}^2+\dfrac{r\sum\sum\sum(\beta\gamma\delta)_{ijk}^2}{(a-1)(b-1)(c-1)}$
	$(\tau\beta\gamma\delta)_{ijkh}$	$\sigma^2+\sigma_{\tau\beta\gamma\delta}^2$
	$\epsilon_{l(ijkh)}$	σ^2(不可估)

14.5.3 带裂区设计

带裂区设计在农业科学中有着广泛的应用, 在工业实验中也有所应用. 最简单情形是我们有两个因子 A 与 B. 因子 A 用于全区, 就像在标准裂区设计中一样. 因子 B 用于带上 (实际上就是另一组全区), 并与刚才因子 A 的全区正交. 图 14.11 给出了 A 和 B 都有三水平时的情形. 注意, A 的各水平与全区混杂, 而因子 B 的各水平与带混杂 (可以看成是 **第 2** 组全区).

图 14.11 的带裂区设计中假设有 r 次重复, A 有 a 个水平, B 有 b 个水平, 则对应的模型是

$$y_{ijk}=\boldsymbol{\mu}+\tau_i+\boldsymbol{\beta}_j+(\boldsymbol{\tau}\beta)_{ij}+\boldsymbol{\gamma}_k+(\tau\boldsymbol{\gamma})_{ik}+(\boldsymbol{\beta\gamma})_{jk}+\boldsymbol{\varepsilon}_{ijk}\quad\begin{cases}i=1,\ 2,\ \cdots,\ r\\ j=1,\ 2,\ \cdots,\ a\\ k=1,\ 2,\ \cdots,\ b\end{cases}$$

其中 $(\tau\beta)_{ij}$ 和 $(\tau\gamma)_{ik}$ 分别是因子 A 和 B 的全区误差, ε_{ijk} 是用于检验 AB 交互作用的 "子区" 误差. 表 14.21 给出了假定 A 和 B 是固定因子而重复是随机因子时的简要方差分析. 重复有时可看作区组.

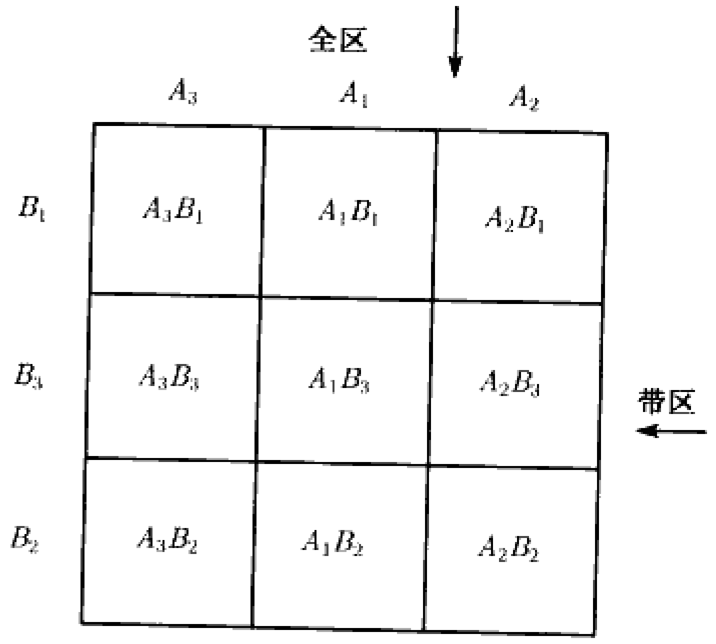


图 14.11　带裂区设计的一次重复 (区组)

表 14.21　带裂区设计的简要方差分析表

方差来源	平方和	自由度	期望均方
重复 (或区组)	$SS_{重复}$	$r-1$	$\sigma_\varepsilon^2+ab\sigma_\tau^2$
A	SS_A	$a-1$	$\sigma_\varepsilon^2+b\sigma_{\tau\beta}^2+\dfrac{rb\sum\beta_j^2}{a-1}$
A 的全区误差	SS_{WP_A}	$(r-1)(a-1)$	$\sigma_\varepsilon^2+b\sigma_{\tau\beta}^2$
B	SS_B	$b-1$	$\sigma_\varepsilon^2+a\sigma_{\tau\gamma}^2+\dfrac{ra\sum\gamma_k^2}{b-1}$
B 的全区误差	SS_{WP_B}	$(r-1)(b-1)$	$\sigma_\varepsilon^2+a\sigma_{\tau\gamma}^2$
AB	SS_{AB}	$(a-1)(b-1)$	$\sigma_\varepsilon^2+\dfrac{r\sum\sum(\tau\beta)_{jk}^2}{(a-1)(b-1)}$
子区误差	SS_{SP}	$(r-1)(a-1)(b-1)$	σ_ε^2
总和	SS_T	$rab-1$	

14.6　思　考　题

*14.1　一位火箭推进剂制造者研究来自 3 种生产过程的火箭推进剂的燃烧率. 从每种过程的产品中随机选取 4 批火箭推进剂. 对每批火箭推进剂进行 3 次燃烧率测试, 其结果如下. 分析这些数据并写出结论.

批次	过程 1				过程 2				过程 3			
	1	2	3	4	1	2	3	4	1	2	3	4
	25	19	15	15	19	23	18	35	14	35	38	25
	30	28	17	16	17	24	21	27	15	21	54	29
	26	20	14	13	14	21	17	25	20	24	50	33

14.2 研究 4 台机器对金属零件进行表面抛光. 进行一项实验, 每台机器由 3 位不同的操作工使用, 从每位操作工中收集两个样品并进行测试. 因为机器位置的缘故, 每台机器为不同的操作工所用, 操作工是随机选取的. 数据如下表所示, 分析这些数据并写出结论.

操作工	机器 1			机器 2			机器 3			机器 4		
	1	2	3	1	2	3	1	2	3	1	2	3
	79	94	46	92	85	76	88	53	46	36	40	62
	62	74	57	99	79	68	75	56	57	53	56	47

*14.3 一位制造工程师研究由 3 台机器生产的某种零件尺寸的变异性. 每台机器有两根转轴, 从每根转轴所加工的零件中随机选取 4 个. 结果如下. 假定机器和转轴是固定因子, 分析这些数据.

转轴	机器 1		机器 2		机器 3	
	1	2	1	2	1	2
	12	8	14	12	14	16
	9	9	15	10	10	15
	11	10	13	11	12	15
	12	8	14	13	11	14

14.4 为了简化生产计划, 一位工业工程师考虑给一类特殊工作指派一个时间标准的可能性, 他相信工作上的差别可忽略不计. 为了了解这种简化是否可能, 随机选取 6 种工作, 每种工作派给分别由 3 位操作工组成的不同小组去做. 每位操作工在一周内的不同时间里完成这种工作两次, 得出的结果如下. 关于这类特殊工作使用共同的时间标准问题, 你的结论是什么? 你会用什么数值作为标准呢?

工作	操作工 1		操作工 2		操作工 3	
1	158.3	159.4	159.2	159.6	158.9	157.8
2	154.6	154.9	157.7	156.8	154.8	156.3
3	162.5	162.6	161.0	158.9	160.5	159.5
4	160.0	158.7	157.5	158.9	161.1	158.5
5	156.3	158.1	158.3	156.9	157.7	156.9
6	163.7	161.0	162.3	160.3	162.6	161.8

*14.5 考虑如图 14.4 所示的三级嵌套设计来研究合金的硬度. 假定合金的化学性质和炉次是固定因子, 铸块是随机的, 用下列数据来分析这一设计.

	合金的化学性质 1						合金的化学性质 2					
炉次	1		2		3		1		2		3	
铸块	1	2	1	2	1	2	1	2	1	2	1	2
	40	27	95	69	65	78	22	23	83	75	61	35
	63	30	67	47	54	45	10	39	62	64	77	42

14.6 使用混合模型的无约束形式, 重新分析思考题 14.5 中的实验. 对你所观测到的有约束和无约束模型分析结果之间的差异进行评论. 可以使用计算机软件.

14.7 在约束混合模型假设下, 假定 A 是固定的, B 和 C 是随机的, 推导出平衡三级嵌套设计的期望均方, 并找出用来估计方差分量的公式.

14.8 在无约束混合模型的假设下，重新考虑思考题 14.7. 可以使用计算机软件. 对有约束和无约束模型的分析和结论之间的差异进行评论.

14.9 如果 3 个因子都是随机的，推导出平衡三级嵌套设计的期望均方. 找出用来估计方差分量的公式.

14.10 证明表 14.1 给出的期望均方.

14.11 **不平衡嵌套设计**. 考虑不平衡二级嵌套设计. 被套因子 B 在因子 A 的第 i 个水平下有 b_i 个水平，在第 (ij) 个单元有 n_{ij} 次重复.

(a) 写出这种情况的最小二乘正规方程组. 解这组正规方程组.

(b) 构造不平衡二级嵌套设计的方差分析表.

(c) 用 (b) 部分的结果分析下列数据：

因子 A	1		2		
因子 B	1	2	1	2	3
	6	−3	5	2	1
	4	1	7	4	0
	8		9	3	−3
			6		

14.12 **不平衡二级嵌套设计的方差分量**. 考虑模型

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{k(ij)} \quad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, b_i \\ k = 1, 2, \cdots, n_{ij} \end{cases}$$

其中 A 和 B 是随机因子，证明

$$\mathrm{E}(MS_A) = \sigma^2 + c_1\sigma_\beta^2 + c_2\sigma_\tau^2, \quad \mathrm{E}(MS_{B(A)}) = \sigma^2 + c_0\sigma_\beta^2, \quad \mathrm{E}(MS_E) = \sigma^2$$

其中

$$c_0 = \frac{N - \sum_{i=1}^{a}\left(\sum_{j=1}^{b_i} n_{ij}^2/n_{i.}\right)}{b-a}, \quad c_1 = \frac{\sum_{i=1}^{a}\left(\sum_{j=1}^{b_i} n_{ij}^2/n_{i.}\right) - \sum_{i=1}^{a}\sum_{j=1}^{b_i} n_{ij}^2/N}{a-1}, \quad c_2 = \frac{N - \frac{\sum_{i=1}^{a} n_{i.}^2}{N}}{a-1}$$

*14.13 一位过程工程师考察制造同一产品的 3 台机器的产量. 每台机器可由两种动力装置来驱动，每台机器有 3 个位置生产产品. 进行一项实验，每台机器都要试验这两种动力装置，每个位置对产品产量进行 3 次测量. 试验以随机顺序进行，结果如下. 假定 3 个因子都是固定的，分析这一实验.

位置	机器 1			机器 2			机器 3		
	1	2	3	1	2	3	1	2	3
动力装置 1	34.1	33.7	36.2	31.1	33.1	32.8	32.9	33.8	33.6
	30.3	34.9	36.8	33.5	34.7	35.1	33.0	33.4	32.8
	31.6	35.0	37.1	34.0	33.9	34.3	33.1	32.8	31.7
动力装置 2	24.3	28.1	25.7	24.1	24.1	26.0	24.2	23.2	24.7
	26.3	29.3	26.1	25.0	25.1	27.1	26.1	27.4	22.0
	27.1	28.6	24.9	26.3	27.9	23.9	25.3	28.0	24.8

14.14 假定在思考题 14.13 中, 有很多动力装置可用, 随机地选取两种进行实验. 求这种情况下的期望均方, 并适当地修改上题的分析.

14.15 在无约束混合模型的假设下, 重新考虑思考题 14.14. 可以使用某个计算机软件包. 对所观测到的约束模型和无约束模型的分析和结论之间的所有差异进行评论.

14.16 一位结构工程师研究从 3 位卖主那里购进的铝合金的强度. 每位卖主提供的合金棒材的标准尺寸是 1.0 英寸、1.50 英寸或 2.0 英寸. 从铸锭制造不同尺寸的棒材的过程涉及不同的锻造工艺, 所以尺寸这一因子是重要的. 还有, 铸锭可来自不同的炉次. 每位卖主对来自 3 个炉次的棒材的各种尺寸都分别提供两件检验样品. 测得强度数据的结果如下. 假定卖主和棒材尺寸是固定的, 炉次是随机的, 使用约束混合模型分析这些数据.

	炉次	卖主 1			卖主 2			卖主 3		
		1	2	3	1	2	3	1	2	3
棒材尺寸	1 英寸	1.230	1.346	1.235	1.301	1.346	1.315	1.247	1.275	1.324
		1.259	1.400	1.206	1.263	1.392	1.320	1.296	1.268	1.315
	$1\frac{1}{2}$ 英寸	1.316	1.329	1.250	1.274	1.384	1.346	1.273	1.260	1.392
		1.300	1.362	1.239	1.268	1.375	1.357	1.264	1.265	1.364
	2 英寸	1.287	1.346	1.273	1.247	1.362	1.336	1.301	1.280	1.319
		1.292	1.382	1.215	1.215	1.328	1.342	1.262	1.271	1.323

14.17 在无约束混合模型的假设下, 重新考虑思考题 14.16. 可以使用计算机软件. 对约束模型和无约束模型的分析和结论之间的差异进行评论.

14.18 假定在思考题 14.16 中, 棒材以多种尺寸购进, 实际做实验时, 随机地选用了 3 种尺寸. 用约束混合模型求这种情况下的期望均方并适当修改上题的分析.

*14.19 钢铁回火是先加热至高于临界温度, 浸一下水, 然后放在空气中冷却. 这道工序会提高钢铁的强度, 去除了杂质颗粒, 并使结构均匀. 进行一项实验来确定温度和热处理时间对回火钢铁强度的效应. 选取 2 种温度和 3 种时间. 进行实验时, 将炉子加热至随机选定的温度并插入 3 块样品. 10 分钟后取走一块样品, 20 分钟后取走第二块, 30 分钟后取走最后一块样品. 然后, 温度改变至另一水平, 重复以上过程. 进行 4 轮实验收集得下列数据. 假定两个因子都是固定的, 分析数据并写出结论.

*14.20 设计一项实验来研究颜料在油漆中的分散性. 研究一种特定颜料的 4 种不同配料. 方法是, 配制好一种配料后, 就用这种配料以 3 种方法 (刷, 喷, 滚) 漆在一块护墙板上. 测量的响应是颜料反射率. 一天一次重复, 共做 3 天, 所得数据如下. 假定配料和油漆方法是固定的, 分析这些数据并写出结论.

轮次	时间 (分)	温度 (°F)		轮次	时间 (分)	温度 (°F)	
		1 500	1 600			1 500	1 600
	10	63	89		10	48	73
1	20	54	91	3	20	74	81
	30	61	62		30	71	69
	10	50	80		10	54	88
2	20	52	72	4	20	48	92
	30	59	69		30	59	64

天	油漆方法	配料			
		1	2	3	4
	1	64.5	66.3	74.1	66.5
1	2	68.3	69.5	73.8	70.0
	3	70.3	73.1	78.0	72.3
	1	65.2	65.0	73.8	64.8
2	2	69.2	70.3	74.5	68.3
	3	71.2	72.8	79.1	71.5
	1	66.2	66.5	72.3	67.7
3	2	69.0	69.0	75.4	68.6
	3	70.8	74.2	80.1	72.4

14.21 重复思考题 14.20, 假定配料是随机的, 油漆方法是固定的.

14.22 考虑例 14.4 描述的裂裂区设计. 假定实验如所描述的那样进行, 所得数据见表 14.22. 分析这些数据并写出结论.

表 14.22 吸收时间实验

		技师								
		1			2			3		
重复 (或区组)	剂量	1	2	3	1	2	3	1	2	3
	糖衣厚度									
	1	95	71	108	96	70	108	95	70	100
1	2	104	82	115	99	84	100	102	81	106
	3	101	85	117	95	83	105	105	84	113
	4	108	85	116	97	85	109	107	87	115
	1	95	78	110	100	72	104	92	69	101
2	2	106	84	109	101	79	102	100	76	104
	3	103	86	116	99	80	108	101	80	109
	4	109	84	110	112	86	109	108	86	113
	1	96	70	107	94	66	100	90	73	98
3	2	105	81	106	100	84	101	97	75	100
	3	106	88	112	104	87	109	100	82	104
	4	113	90	117	121	90	117	110	91	112
	1	90	68	109	98	68	106	98	72	101
4	2	100	84	112	102	81	103	102	78	105
	3	102	85	115	100	85	110	105	80	110
	4	114	88	118	118	85	116	110	95	120

14.23 再做思考题 14.22, 假定剂量是随机选取的, 并使用约束混合模型.

14.24 设在思考题 14.22 中用了 4 位技师. 假定所有因子是固定的, 要求一个适合于检验技师之间的差异的自由度, 应该做多少个区组的试验?

14.25 考虑例 14.4 描述的实验, 指出做试验时怎样确定处理组合的顺序, 如果此实验按以下设计考虑: (a) 裂裂区设计, (b) 裂区设计, (c) 随机化区组析因设计, (d) 完全随机化析因设计.

14.26 *Quality Engineering* ("Quality Quandries: Two-Level Factorials Run as Split-Plot Experi-

ments" Bisgaard 等人, Vol.8, No.4, pp.705~708, 1996) 上的一篇文章描述了一个纸浆制造过程中的 2^5 析因实验, 目的是使纸张对墨水更敏感. 其中有 4 个因子 ($A \sim D$) 在试验间难以改变, 因此实验者把反应器按照这些因子的高低水平设置在 8 组条件下, 然后对两种纸张类型 (因子 E) 同时处理. 纸张样品在反应器里的放置 (右对左) 是随意的. 这样就得到了一个裂区设计, $A \sim D$ 是全区因子, E 是子区因子. 实验数据如下. 分析这些实验数据并写出结论.

标准次序	实验号	$A=$压力	$B=$功率	$C=$气流	$D=$气体类型	$E=$纸张类型	y 接触角度	标准次序	实验号	$A=$压力	$B=$功率	$C=$气流	$D=$气体类型	$E=$纸张类型	y 接触角度
5	1	−1	−1	+1	O_2	E1	37.6	16	17	+1	+1	+1	S_iCl_4	E1	49.5
21	2	−1	−1	+1	O_2	E2	43.5	32	18	+1	+1	+1	S_iCl_4	E2	48.2
2	3	+1	−1	−1	O_2	E1	41.2	9	19	−1	−1	−1	S_iCl_4	E1	5.0
18	4	+1	−1	−1	O_2	E2	38.2	25	20	−1	−1	−1	S_iCl_4	E2	18.1
10	5	+1	−1	−1	S_iCl_4	E1	56.8	15	21	−1	+1	+1	S_iCl_4	E1	11.3
26	6	+1	−1	−1	S_iCl_4	E2	56.2	31	22	−1	+1	+1	S_iCl_4	E2	23.9
14	7	+1	−1	+1	S_iCl_4	E1	47.5	1	23	−1	−1	−1	O_2	E1	48.6
30	8	+1	−1	+1	S_iCl_4	E2	43.2	17	24	−1	−1	−1	O_2	E2	57.0
11	9	−1	+1	−1	S_iCl_4	E1	25.6	8	25	+1	+1	+1	O_2	E1	48.7
27	10	−1	+1	−1	S_iCl_4	E2	33.0	24	26	+1	+1	+1	O_2	E2	44.4
3	11	−1	+1	−1	O_2	E1	55.8	7	27	−1	+1	+1	O_2	E1	47.2
19	12	−1	+1	−1	O_2	E2	62.9	23	28	−1	+1	+1	O_2	E2	54.6
13	13	−1	−1	+1	S_iCl_4	E1	13.3	4	29	+1	+1	−1	O_2	E1	53.5
29	14	−1	−1	+1	S_iCl_4	E2	23.7	20	30	+1	+1	−1	O_2	E2	51.3
6	15	+1	−1	+1	O_2	E1	47.2	12	31	+1	+1	−1	S_iCl_4	E1	41.8
22	16	+1	−1	+1	O_2	E2	44.8	28	32	+1	+1	−1	S_iCl_4	E2	37.8

14.27 重新考虑思考题 14.26. 这是一个很大的实验, 所以假设实验者用了一个 2^{5-1} 设计来代替. 用主分式建立有裂区结构的 2^{5-1} 设计, 响应值从完全析因情形中选取得到. 试分析数据并写出结果. 这些结果与思考题 14.26 中的结果一致吗?

第 15 章　其他设计与分析论题

本章纲要

15.1　非正态响应与变换
　15.1.1　选择一个变换：Box-Cox 方法
　15.1.2　广义线性模型
15.2　析因设计中的不平衡数据
　15.2.1　成比例的数据：简单的情况
　15.2.2　近似方法
　15.2.3　精确方法
15.3　协方差分析
　15.3.1　方法说明
　15.3.2　计算机输出
　15.3.3　用一般线性回归显著性检验进行推导
　15.3.4　含协变量的析因实验
15.4　重复测量
第 15 章补充材料
S15.1　变换的形式
S15.2　Box-Cox 方法中 λ 的选择
S15.3　广义线性模型
　S15.3.1　二元响应变量的模型
　S15.3.2　逻辑斯谛回归模型中的参数估计
　S15.3.3　逻辑斯谛回归模型中的参数解释
　S15.3.4　模型参数的假设检验
　S15.3.5　泊松回归
　S15.3.6　广义线性模型
　S15.3.7　联系函数与线性预报元
　S15.3.8　广义线性模型中的参数估计
　S15.3.9　广义线性模型的预测与估计
　S15.3.10　广义线性模型的残差分析
S15.4　析因设计中的不平衡数据
　S15.4.1　回归模型法
　S15.4.2　第 3 型分析
　S15.4.3　第 1 型、第 2 型、第 3 型及第 4 型平方和
　S15.4.4　使用均值模型进行不平衡数据分析
S15.5　计算机实验

统计设计实验的论题是非常广泛的. 前面各章已经初步介绍了许多基本概念和方法, 不过有时我们也只能给出一个概要. 例如, 像响应曲面方法、混料实验、方差分量估计以及最优设计等论题其内容都多到可以单独出书. 本章将概要介绍其他几个论题, 实验者会发现它们是有用的.

15.1　非正态响应与变换

15.1.1　选择一个变换：Box-Cox 方法

3.4.3 节讨论了设计实验中响应变量的非常数方差问题, 并指出这违背了标准方差分析的假定. 方差不等的问题在实际中经常出现, 而且通常与非正态响应变量有关. 譬如说, 对次品或粒子的计数, 诸如产率或次品率的比例数据, 或者分布有偏倚 (分布的某一尾比另一尾更长些) 的响应变量. 我们要介绍的**响应变量变换**是用于稳定响应方差的合适方法. 下面讨论选择变换形式的两种方法 —— 经验图形法和试错法 (即实验者只尝试一种或几种变换, 并从中选出能产生最满意的拟合响应残差图的那种变换).

一般而言, 变换有 3 种用途：稳定响应方差, 使响应变量的分布近似于正态分布, 增强模型对数据的拟合度. 最后的那种用途包括通过去除交互作用来简化模型. 有时一个变换可以相当有

效地同时实现上述多种目的.

幂变换族$y* = y^{\lambda}$ 是非常有用的, 其中 λ 是变换的待定参数 (如, $\lambda = \frac{1}{2}$ 是指用原响应的平方根). Box and Cox(1964) 说明了如何同时估计变换参数 λ 与其他模型参数 (总体均值与处理效应). 该方法的理论依据是最大似然估计. 实际计算步骤包括, 对不同的 λ 值, 关于

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda}-1}{\lambda \dot{y}^{\lambda-1}} & \lambda \neq 0 \\ \dot{y}\ln y & \lambda = 0 \end{cases} \tag{15.1}$$

进行标准的方差分析, 其中 $\dot{y} = \ln^{-1}[(1/n)\sum \ln y]$ 是观测的几何均值. λ 的最大似然估计是使得误差平方和 $SS_E(\lambda)$ 最小的值. 通常确定 λ 的步骤如下: 先画出 $SS_E(\lambda)$ 关于 λ 的图像, 然后读出图中使 $SS_E(\lambda)$ 最小的 λ 值. 一般 10 到 20 个 λ 值足以估计出最优值. 若要更精确的 λ 的估计, 可以用更精细的网格值接着再做.

注意, 我们不能通过直接比较关于 y^{λ} 的方差分析的误差平方和来选择 λ, 这是因为对于不同的 λ 值误差平方和的测量尺度也是不同的. 而且, 当 $\lambda = 0$ 时 y 会有问题, 即当 λ 接近 0 时, y^{λ} 接近 1. 所以, $\lambda = 0$ 时, 所有响应值都是常数. (15.1) 式中的项 $(y^{\lambda}-1)/\lambda$ 消除了这种问题, 因为当 λ 趋近 0 时, $(y^{\lambda}-1)/\lambda$ 趋近于极限 $\ln y$. (15.1) 式中的除式 $\dot{y}^{\lambda-1}$ 对响应建立了新尺度, 使得误差平方和可直接比较.

使用 Box-Cox 方法时, 建议实验者使用 λ 值的简单选择. 因为 $\lambda = 0.5$ 与 $\lambda = 0.58$ 间的实际差别应该很小, 但是平方根变换 ($\lambda = 0.5$) 更易于解释. 显然, λ 值接近于 1 时表明没必要作任何变换.

一旦用 Box-Cox 方法选出了 λ 值, 实验者可用 y^{λ} 作为响应进行数据分析, 当然 $\lambda = 0$ 时, 可以用 $\ln y$. 用 $y^{(\lambda)}$ 作为实际响应是非常好的, 尽管模型参数估计将有不同的尺度, 且与使用 y^{λ}(或 $\ln y$) 所得的结果相比有坐标平移.

λ 的近似 $100(1-\alpha)\%$ 置信区间可通过计算

$$SS^* = SS_E(\lambda)\left(1 + \frac{t^2_{\alpha/2,\nu}}{\nu}\right) \tag{15.2}$$

来得到, 其中 ν 是自由度. 在 $SS_E(\lambda)$ 关于 λ 的图像中, 在 SS^* 的高度处作平行于 λ 轴的直线. 然后定出 SS^* 直线与曲线 $SS_E(\lambda)$ 交点在 λ 轴上的坐标, 可以从图上直接读出 λ 的置信限. 若该置信区间包含了 $\lambda = 1$, 则表明 (如上所说) 这些数据没有必要进行变换.

例 15.1 用原先在例 3.5 给出的洪峰流量数据来说明 Box-Cox 方法. 这是个单因子实验 (原始数据见表 3.7). 用 (15.1) 式, 算出对应于不同 λ 值的 $SS_E(\lambda)$ 值:

λ	−1.00	−0.50	−0.25	0.00	0.25	0.50	0.75	1.00	1.25	1.50
$SS_E(\lambda)$	7 922.11	687.10	232.52	91.96	46.99	35.42	40.61	62.08	109.82	208.12

最小值附近的取值图像见图 15.1, 由此可见, $\lambda \approx 0.52$ 时给出了最小值, 约为 $SS_E(\lambda) = 35.00$. 以下通过计算 (15.2) 式的量 $SS*$, 得到 λ 的近似 95%置信区间:

$$SS^* = SS_E(\lambda)\left(1 + \frac{t^2_{0.025,20}}{20}\right) = 35.00\left[1 + \frac{(2.086)^2}{20}\right] = 42.61$$

在图 15.1 中画出 SS^*, 读出直线与曲线交点的 λ 坐标, 得到 λ 的置信下限与上限分别为 $\lambda^- = 0.27$ 和 $\lambda^+ = 0.77$. 因为置信限内不包括值 1, 所以需要使用变换, 而且实际上使用平方根变换 ($\lambda = 0.50$) 更加合理.

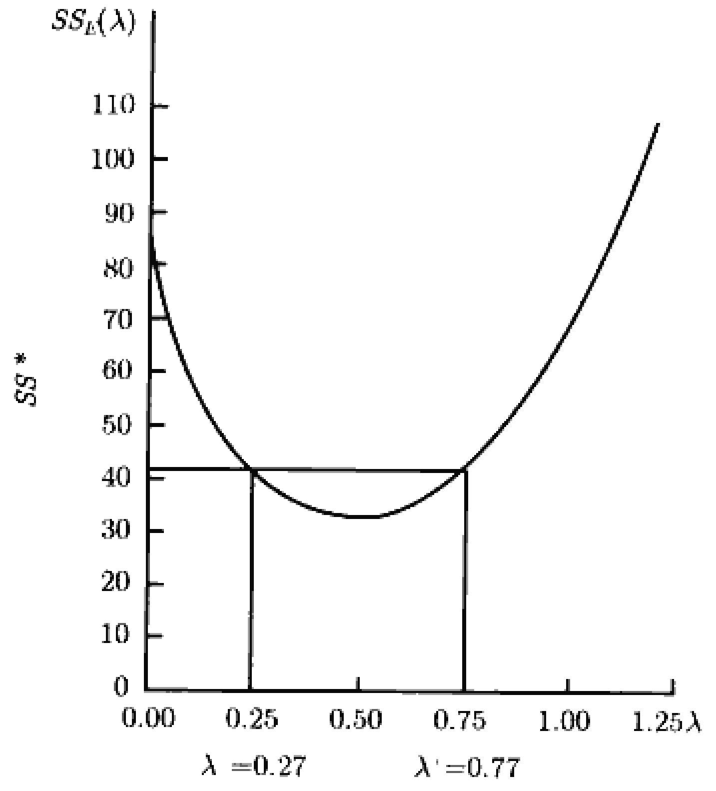


图 15.1 例 15.1 中 $SS_E(\lambda)$ 与 λ 的关系图

一些计算机程序都含有 Box-Cox 方法, 用于选择幂变换族. 图 15.2 给出了 Design-Expert 关于洪峰流量数据的该方法输出. 此计算与汇总在例 15.1 中的手算结果非常吻合. 注意图 15.2 中图像的纵轴是 $\ln[SS_E(\lambda)]$.

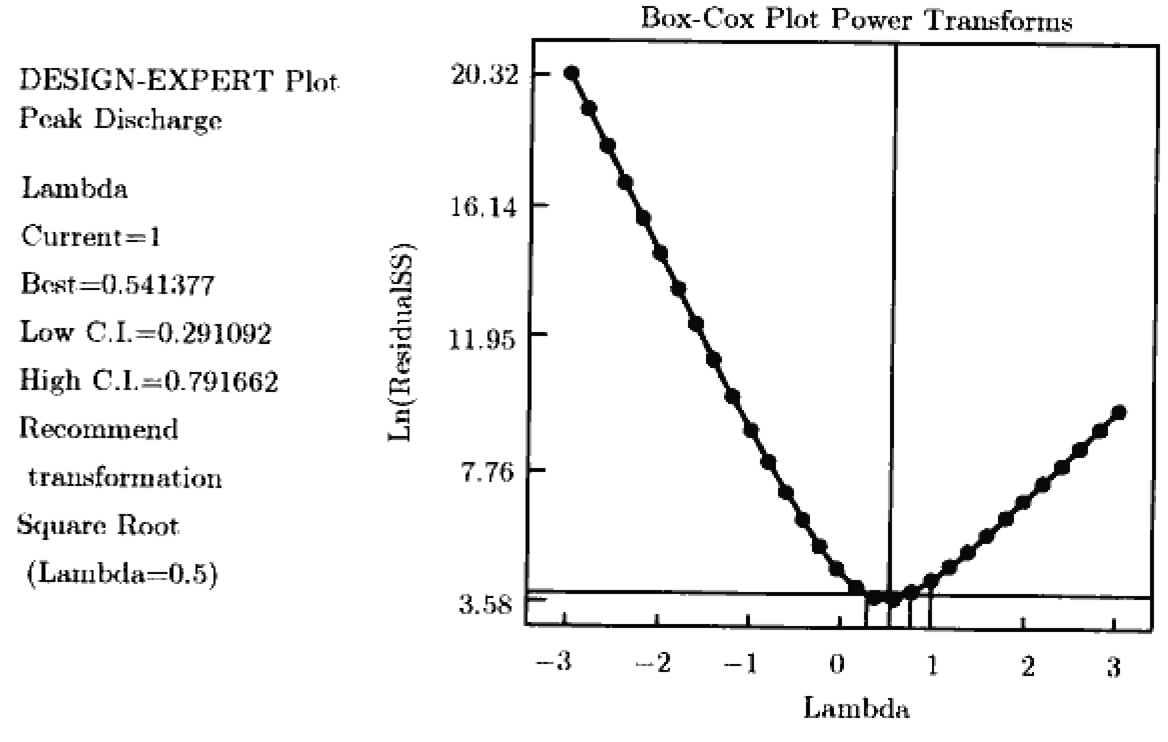


图 15.2 Design-Expert 的 Box-Cox 方法输出

15.1.2 广义线性模型

数据变换常常是处理非正态响应以及相应的不等方差问题的非常有效的方法. 正如 15.1.1

节所讲的, Box-Cox 方法是选择变换形式的简单有效的方法. 但是, 使用数据变换也存在着一些问题.

一个问题是, 实验者可能不习惯使用变换尺度后的响应变量. 譬如, 他对次品数量感兴趣, 但对次品数量的平方根没兴趣, 或者对电阻率感兴趣而对电阻率的对数没兴趣. 相反, 若变换确实很成功, 且改进了对响应的分析与相关模型, 实验者通常会很快采用这种新的测量.

一个更严重的问题是, 变换可能导致在实验者感兴趣的设计因子空间的某个部分出现响应变量的**无意义值**. 例如, 在一个关于半导体晶片上观测到的瑕疵数的实验中, 假定我们已经使用了平方根变换, 结果在某些感兴趣的区域, 瑕疵数的预测平方根是负数. 当观测到的实际瑕疵数很小时, 这是有可能发生的. 因此, 实验的模型恰在期待它有较好预测性能的区域中产生了不可靠的预测值.

最后, 如 15.1.1 节所指出的, 我们经常采用变换来稳定方差、改进正态性并简化模型. 但不能保证一个变换就能同时完成所有这些目的.

不同于这种先做数据变换然后对变换后的响应进行标准最小二乘分析的典型方法, 另一种方法是采用**广义线性模型(GLM)**. 这是 Nelder and Wedderburn(1972) 提出的方法, 它本质上统一了含有正态和非正态响应的线性与非线性模型. McCullagh and Nelder(1989) 及 Myers, Montgomery, and Vining(2002) 给出了广义线性模型的详细介绍, 而且 Myers and Montgomery(1997) 给出过指南. 本章的补充材料里也给出了更多的细节. 我们将在下面给出这些概念的概要, 并用 3 个简短的例子来说明它们.

广义线性模型本质上是一个**回归模型**(实验设计模型也是回归模型). 与所有的回归模型一样, 它由随机项 (通常称为误差项) 以及一个关于设计因子 (那些 x) 和一些未知参数 (那些 β) 的函数所组成. 标准正态线性回归模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{15.3}$$

其中假定误差项 ε 服从均值为零且方差为常数的正态分布, 而且响应变量 y 的均值为

$$\mathrm{E}(y) = \mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k = \boldsymbol{x}'\boldsymbol{\beta} \tag{15.4}$$

(15.4) 式的 $\boldsymbol{x}'\boldsymbol{\beta}$ 部分称为**线性预报元**. 广义线性模型把 (15.3) 式作为它的一种特殊情况.

在一个广义线性模型中, 响应变量可以是**指数型分布族**中的任一分布. 该分布族包括**正态**、**泊松**、**二项**、**指数**和**伽玛**分布, 所以指数型分布族是非常广泛的且非常灵活的分布的集合, 能用于很多实验场合. 而且, 响应均值与线性预报元的关系是通过**联系函数** (link function) 确定的.

$$g(\mu) = \boldsymbol{x}'\boldsymbol{\beta} \tag{15.5}$$

表示响应均值的回归模型表示为

$$\mathrm{E}(y) = \mu = g^{-1}(\boldsymbol{x}'\boldsymbol{\beta}) \tag{15.6}$$

例如, (15.3) 式中普通线性回归模型的联系函数称为**恒等联系**, 这因为 $\mu = g^{-1}(\boldsymbol{x}'\boldsymbol{\beta}) = \boldsymbol{x}'\boldsymbol{\beta}$. 作为另一个例子, **对数联系**

$$\ln(\mu) = \boldsymbol{x}'\boldsymbol{\beta} \tag{15.7}$$

得到模型

$$\mu = e^{\boldsymbol{x}'\boldsymbol{\beta}} \tag{15.8}$$

对数联系常用于计数数据 (泊松响应) 以及具有右侧长尾分布的连续型响应 (指数分布或伽玛分布). 另一种常用于二项分布数据的重要联系函数是**Logit 联系**

$$\ln\left(\frac{\mu}{1-\mu}\right) = \boldsymbol{x}'\boldsymbol{\beta} \tag{15.9}$$

选择使用这种联系函数得到模型

$$\mu = \frac{1}{1+e^{-\boldsymbol{x}'\boldsymbol{\beta}}} \tag{15.10}$$

联系函数可能有多个选择, 但它必须是单调可微的. 而且要注意, 在广义线性模型中, 响应变量的方差不一定是常量, 它可以是均值的函数 (及通过联系函数也是预报元的函数). 例如, 若响应是泊松分布的, 则响应的方差正好等于均值.

为了在实际中使用广义线性模型, 实验者必须指定响应分布和联系函数. 然后, 用最大似然法进行模型拟合与参数估计, 对于指数型分布族来讲, 就变成了迭代加权最小二乘形式. 对于正态响应变量的普通线性模型或实验设计模型, 这归结为标准的最小二乘法. 用类似于正态理论数据的方差分析方法, 可以对广义线性模型进行推断或诊断性检查. 细节与例子参见 Myers and Montgomery(1997) 以及 Myers, Montgomery and Vining(2002). 可以很好地支持广义线性模型的两个软件包是 SAS(PROC GENMOD) 和 S-PLUS.

例 15.2　一家消费品生产公司研究影响顾客用优惠券兑换该公司个人护理产品的几率的因子. 一个 2^3 析因实验用于研究下列变量: A= 优惠券的价值 (低, 高), B= 优惠券的有效期, C= 使用的方便程度 (容易, 困难). 2^3 设计的 8 个单元的每个单元中都随机安排了 1 000 人, 响应变量是兑换的优惠券的数量. 实验结果列在表 15.1 中.

表 15.1　优惠券兑换实验的设计与数据

A	B	C	优惠券兑换数
−	−	−	200
+	−	−	250
−	+	−	265
+	+	−	347
−	−	+	210
+	−	+	286
−	+	+	271
+	+	+	326

设计的每个单元中的响应值可以看作 1 000 重贝努利试验中的成功次数, 所以该响应的一个合理模型是, 具有二项响应分布和 logit 联系函数的广义线性模型. 这种特殊形式的广义线性模型通常称为**逻辑斯谛回归** (logistic regression).

Minitab 能拟合逻辑斯谛回归模型. 实验者决定拟合只包含主效应和二因子交互作用的模型. 因此, 响应均值的模型为

$$\mathrm{E}(y)=\frac{1}{1+\exp\left[-\left(\beta_0+\sum\limits_{i=1}^{3}\beta_i x_i+\sum\limits_{i<}^{2}\sum\limits_{j=2}^{3}\beta_{ij}x_i x_j\right)\right]}$$

表 15.2 给出了针对表 15.1 中数据的 Minitab 的部分输出. 表的上半部分拟合了全模型, 包含了所有 3 个主效应和 3 个二因子交互作用. 可看到输出包含模型系数的估计及它们的标准误. 在零假设回归系数等于零成立时, 模型系数的估计与其标准误的比值 (t 型比值) 有近似正态分布. 因此, 可用比值 $Z=\hat{\beta}/se(\hat{\beta})$ 检验各主效应或交互作用项是否显著. 这里 "近似" 意味着样本量很大. 现在这里的样本量并不很大, 我们要小心解释这些 t 型比值的 P 值, 但是这些统计量可用作数据分析的前导. 表中的 P 值表明截距项、A 与 B 的主效应以及 BC 交互作用是显著的.

表中拟合优度部分给出用来衡量模型总体适合程度的 3 个不同的检验统计量 (Pearson, Deviance 和 Hosmer-Lemeshow). 这些拟合优度统计量的 P 值都很大, 这表明模型是令人满意的. 所拟合的模型是

$$\begin{aligned}\hat{y}&=\frac{1}{1+\exp[-(-1.01+0.169x_1+0.169x_2+0.023x_3-0.041x_2x_3)]}\\&=\frac{1}{1+\exp(+1.01-0.169x_1-0.169x_2-0.023x_3+0.041x_2x_3)}\end{aligned}$$

因为效应 C 和交互作用 BC 很小, 这些项好像可以从模型中剔除且不会有大的影响.

Minitab 对每个回归模型系数都给出了**优势比** (odd ratio). 优势比直接从 (15.9) 式的 logit 联系函数得到, 非常类似于标准二水平设计中的因子效应估计. 对于因子 A, 它可解释为 A 取高的值 ($x_1=+1$) 时兑换优惠券的几率与 A 取 $x_1=0$ 时兑换优惠券的几率的比值. 算得的优势比为 $e^{\hat{\beta}_1}=e^{0.168\ 765}=1.18$. 量 $e^{2\hat{\beta}_1}=e^{2(0.168\ 765)}=1.40$ 是 A 取高的值 ($x_1=+1$) 时兑换优惠券的几率与 A 取低的值 ($x_1=-1$) 时兑换优惠券的几率的比值. 因此, 高值的优惠券增加了 40%的兑换优惠券的几率.

逻辑斯谛回归应用很广泛, 它可能是应用最广的广义线性模型. 它在关于剂量–响应研究 (设计因子是特定的治疗处理的剂量, 而响应为病人是否被成功治好) 的生物医学领域有着广泛的应用. 许多可靠性工程实验包含二元 (成–败) 数据, 例如当单位产品或部件承受压力或负荷, 响应是该单位是否失效.

例 15.3 护栅板缺陷实验

思考题 8.30 介绍了一个实验, 用来研究 9 个因子对可开式模压护栅板的影响. Bisdaard 和 Fuller (1994–1995) 对该数据进行了有趣且有用的分析, 来说明实验设计中数据变换的作用. 如思考题 8.30 中 (f) 部分所看到的, 他们用修改的平方根变换得到模型:

$$\widehat{(\sqrt{y}+\sqrt{y+1})/2}=2.513-0.996x_4-1.21x_6-0.772x_2x_7$$

一般情况下, x 表示的是规范后的设计因子. 该变换很好地稳定了次品数的方差. 表 15.3 的前两组给出此模型的一些信息. 在 "变换后" 标题下, 第一列是响应的预测值. 注意那里有两个负的预测值. 标题 "未变换" 下给出未变换的预测值, 以及在 16 个设计点的各点处的响应均值的 95%置信区间. 因为有一些负的预测值或置信下限, 我们无法算出该表中的所有单元中的值.

响应实际上是次品数的平方根. 负的预测值显然不合逻辑. 注意, 这种情况在观测到的数值很小时发生. 若在这种区域中用模型预测性能表现是很重要的, 则此模型也许不值得信赖. 这不应被当作对实验者或 Bisdaard 和 Fuller 的分析的批评. 这是一个极其成功的筛选实验, 它能清楚区分出重要的工序变量. 预测并非原先的目的, 也不是 Bisdaard 和 Fuller 所做分析的目的.

表 15.2 优惠券兑换实验的 Minitab 输出

二项逻辑斯谛回归：全模型

Link Function: Logit

Response Information

Variable	Value	Count
C5	Success	2155
	Failure	5845
C6	Total	8000

Logistic Regression Table

Predictor	Coef	SE Coef	z	P	odds Ratio	95% Lower	CI Upper
Constant	-1.01154	0.0255150	-39.65	0.000			
A	0.169208	0.0255092	6.63	0.000	1.18	1.13	1.25
B	0.169622	0.0255150	6.65	0.000	1.18	1.13	1.25
C	0.0233173	0.0255099	0.91	0.361	1.02	0.97	1.08
A*B	-0.0062854	0.0255122	-0.25	0.805	0.99	0.95	1.04
A*C	-0.0027726	0.0254324	-0.11	0.913	1.00	0.95	1.05
B*C	-0.0410198	0.0254339	-1.61	0.107	0.96	0.91	1.01

Log-Likelihood = -4615.310

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1.46458	1	0.226
Deviance	1.46451	1	0.226
Hosmer-Lemeshow	1.46458	6	0.962

二项逻辑斯谛回归：简化模型，涉及 A, B, C, BC

Link Function: Logit

Response Information

Variable	Value	Count
C5	Success	2155
	Failure	5845
C6	Total	8000

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	odds Ration	95% Lower	CI Upper
Constant	-1.01142	0.0255076	-39.65	0.000			
A	0.168675	0.0254235	6.63	0.000	1.18	1.13	1.24
B	0.169116	0.0254321	6.65	0.000	1.18	1.13	1.24
C	0.0230825	0.0254306	0.91	0.364	1.02	0.97	1.08
B*C	-0.0409711	0.0254307	-1.61	0.107	0.96	0.91	1.01

Log-Likelihood = -4615.346

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1.53593	3	0.674
Deviance	1.53602	3	0.674
Hosmer-Lemeshow	1.53593	6	0.957

表 15.3 可开式模压护栅板实验的最小二乘模型与广义线性模型分析

观测	基于 Freeman 和 Tukey 修改的平方根数据变换的最小二乘方法 变换后 预测值	变换后 95%置信区间	未变换 预测值	未变换 95%置信区间	广义线性模型 (泊松响应, 对数联系函数) 预测值	广义线性模型 95%置信区间	95%置信区间的长度 最小二乘	95%置信区间的长度 广义线性模型
1	5.50	(4.14, 6.85)	29.70	(16.65, 46.41)	51.26	(42.45, 61.90)	29.76	19.45
2	3.95	(2.60, 5.31)	15.12	(6.25, 27.65)	11.74	(8.14, 16.94)	21.39	8.80
3	1.52	(0.17, 2.88)	1.84	(1.69, 7.78)	1.12	(0.60, 2.08)	6.09	1.47
4	3.07	(1.71, 4.42)	8.91	(2.45, 19.04)	4.88	(2.87, 8.32)	16.59	5.45
5	1.52	(0.17, 2.88)	1.84	(1.69, 7.78)	1.12	(0.60, 2.08)	6.09	1.47
6	3.07	(1.71, 4.42)	8.91	(2.45, 19.04)	4.88	(2.87, 8.32)	16.59	5.45
7	5.50	(4.14, 6.85)	29.70	(16.65, 46.41)	51.26	(42.45, 61.90)	29.76	19.45
8	3.95	(2.60, 5.31)	15.12	(6.25, 27.65)	11.74	(8.14, 16.94)	21.39	8.80
9	1.08	(−0.28, 2.43)	0.71	(*, 5.41)	0.81	(0.42, 1.56)	*	1.13
10	−0.47	(−1.82, 0.89)	*	(*, 0.36)	0.19	(0.09, 0.38)	*	0.29
11	1.96	(0.61, 3.31)	3.36	(0.04, 10.49)	1.96	(1.16, 3.30)	10.45	2.14
12	3.50	(2.15, 4.86)	11.78	(4.13, 23.10)	8.54	(5.62, 12.98)	18.96	7.35
13	1.96	(0.61, 3.31)	3.36	(0.04, 10.49)	1.96	(1.16, 3.30)	10.45	2.14
14	3.50	(2.15, 4.86)	11.78	(4.13, 23.10)	8.54	(5.62, 12.98)	18.97	7.35
15	1.08	(−0.28, 2.43)	0.71	(*, 5.41)	0.81	(0.42, 1.56)	*	1.13
16	−0.47	(−1.82, 0.89)	*	(*, 0.36)	0.19	(0.09, 0.38)	*	0.29

不过, 若得到预测模型是重要的, 则广义线性模型也许是不同于变换法的一种好方法. Myers 和 Montgomery 用对数联系函数 [(15.7) 式] 和泊松响应, 拟合与 Bisdaard 和 Fuller 所做的相同的线性预报元. 得到模型

$$\hat{y} = \mathrm{e}^{(1.128-0.896x_4-1.176x_6-0.737x_2x_7)}$$

表 15.3 中的第 3 组包括由此模型得到的预测值, 以及设计各点处响应均值的 95%置信区间 (用 SAS PROC GENMOD 算得). 没有负的预测值 (选取联系函数来保证) 和负的置信下限. 表中最后一组比较了未变换响应的 95%置信区间的长度与广义线性模型的 95%置信区间的长度. 可以看到广义线性模型得到的置信区间长度都比最小二乘法的短. 这强烈表明广义线性模型法解释了更多的变异, 相对于变换法能得到更好的模型.

例 15.4 精纺毛纱实验

表 15.4 给出了一个用于研究精纺毛纱的 3^3 析因设计. 该实验是 Box and Draper(1987) 描述的. 响应是到失效时的圈数. 这类可靠性数据是典型的非负、连续且通常服从右侧长尾的分布.

数据原来是用标准 (最小二乘) 方法分析的, 所以需要数据变换来稳定方差. 从总体模型拟合以及令人满意的残差图来看, 根据到失效时的圈数的自然对数数据可以得到一个合适的模型. 该模型为

$$\ln\hat{y} = 6.33 + 0.82x_1 - 0.63x_2 - 0.38x_3$$

或用原来的响应 (到失效时的圈数) 表示为

$$\hat{y} = \mathrm{e}^{6.33+0.82x_1-0.63x_2-0.38x_3}$$

该实验也可以用广义线性模型并选取伽玛分布的响应和对数联系函数来分析. 采用与对数变换响应的

最小二乘分析时一样的模型. 这样得到的模型为

$$\hat{y} = e^{6.35+0.84x_1-0.63x_2-0.39x_3}$$

表 15.5 列出了由最小二乘模型和广义线性模型得到的预测值, 以及在该设计 27 个点的各点处均值的 95%置信区间. 对置信区间长度的比较表明, 广义线性模型好像比最小二乘模型预测得更好.

表 15.4 精纺毛纱实验

试验	x_1	x_2	x_3	到失效时的圈数	到失效时的圈数的自然对数
1	−1	−1	−1	674	6.51
2	−1	−1	0	370	5.91
3	−1	−1	1	292	5.68
4	−1	0	−1	338	5.82
5	−1	0	0	266	5.58
6	−1	0	1	210	5.35
7	−1	1	−1	170	5.14
8	−1	1	0	118	4.77
9	−1	1	1	90	4.50
10	0	−1	−1	1 414	7.25
11	0	−1	0	1 198	7.09
12	0	−1	1	634	6.45
13	0	0	−1	1 022	6.93
14	0	0	0	620	6.43
15	0	0	1	438	6.08
16	0	1	−1	442	6.09
17	0	1	0	332	5.81
18	0	1	1	220	5.39
19	1	−1	−1	3 636	8.20
20	1	−1	0	3 184	8.07
21	1	−1	1	2 000	7.60
22	1	0	−1	1 568	7.36
23	1	0	0	1 070	6.98
24	1	0	1	566	6.34
25	1	1	−1	1 140	7.04
26	1	1	0	884	6.78
27	1	1	1	360	5.89

表 15.5 精纺毛纱实验的最小二乘模型与广义线性模型分析

	基于对数变换采用最小二乘方法							
	变换后		未变换		广义线性模型		95%置信区间的长度	
观测	预测值	95%置信区间	预测值	95%置信区间	预测值	95%置信区间	最小二乘	广义线性模型
1	2.83	(2.76, 2.91)	682.50	(573.85, 811.52)	680.52	(583.83, 793.22)	237.67	209.39
2	2.66	(2.60, 2.73)	460.26	(397.01, 533.46)	463.00	(407.05, 526.64)	136.45	119.59
3	2.49	(2.42, 2.57)	310.38	(260.98, 369.06)	315.01	(271.49, 365.49)	108.09	94.00
4	2.56	(2.50, 2.62)	363.25	(313.33, 421.11)	361.96	(317.75, 412.33)	107.79	94.58
5	2.39	(2.34, 2.44)	244.96	(217.92, 275.30)	246.26	(222.55, 272.51)	57.37	49.96
6	2.22	(2.15, 2.28)	165.20	(142.50, 191.47)	167.55	(147.67, 190.10)	48.97	42.42

(续)

观测	基于对数变换采用最小二乘方法 变换后 预测值	95%置信区间	未变换 预测值	95%置信区间	广义线性模型 预测值	95%置信区间	95%置信区间的长度 最小二乘	广义线性模型
7	2.29	(2.21, 2.36)	193.33	(162.55, 229.93)	192.52	(165.69, 223.70)	67.38	58.01
8	2.12	(2.05, 2.18)	130.38	(112.46, 151.15)	130.98	(115.43, 148.64)	38.69	33.22
9	1.94	(1.87, 2.02)	87.92	(73.93, 104.54)	89.12	(76.87, 103.32)	30.62	26.45
10	3.20	(3.13, 3.26)	1 569.28	(1 353.94, 1 819.28)	1 580.00	(1 390.00, 1 797.00)	465.34	407.00
11	3.02	(2.97, 3.08)	1 058.28	(941.67, 1 189.60)	1 075.00	(972.52, 1 189.00)	247.92	216.48
12	2.85	(2.79, 2.92)	713.67	(615.60, 827.37)	731.50	(644.35,830.44)	211.77	186.09
13	2.92	(2.87, 2.97)	835.41	(743.19, 938.86)	840.54	(759.65, 930.04)	195.67	170.39
14	2.75	(2.72, 2.78)	563.25	(523.24, 606.46)	571.87	(536.67, 609.38)	83.22	72.70
15	2.58	(2.53, 2.63)	379.84	(337.99, 426.97)	389.08	(351.64, 430.51)	88.99	78.87
16	2.65	(2.58, 2.71)	444.63	(383.53, 515.35)	447.07	(393.81, 507.54)	131.82	113.74
17	2.48	(2.43, 2.53)	299.85	(266.75, 336.98)	304.17	(275.13, 336.28)	70.23	61.15
18	2.31	(2.24, 2.37)	202.16	(174.42, 234.37)	206.95	(182.03, 235.27)	59.95	53.23
19	3.56	(3.48, 3.63)	3 609.11	(3 034.59, 4 292.40)	3 670.00	(3 165.00, 4 254.00)	1 257.81	1 089.00
20	3.39	(3.32, 3.45)	2 433.88	(2 099.42, 2 821.63)	2 497.00	(2 200.00, 2 833.00)	722.21	633.00
21	3.22	(3.14, 3.29)	1 641.35	(1 380.07, 1 951.64)	1 699.00	(1 462.00, 1 974.00)	571.57	512.00
22	3.28	(3.22, 3.35)	1 920.88	(1 656.91, 2 226.90)	1 952.00	(1 720.00, 2 215.00)	569.98	495.00
23	3.11	(3.06, 3.16)	1 295.39	(1 152.66, 1 455.79)	1 328.00	(1 200.00, 1 470.00)	303.14	270.00
24	2.94	(2.88, 3.01)	873.57	(753.53, 1 012.74)	903.51	(793.15, 1 029.00)	259.22	235.85
25	3.01	(2.93, 3.08)	1 022.35	(859.81, 1 215.91)	1 038.00	(894.79, 1 205.00)	356.10	310.21
26	2.84	(2.77, 2.90)	689.45	(594.70, 799.28)	706.34	(620.99, 803.43)	204.58	182.44
27	2.67	(2.59, 2.74)	464.94	(390.93, 552.97)	480.57	(412.29, 560.15)	162.04	147.86

广义线性模型已在生药的研究与开发中得到了广泛的应用. 随着更多的软件包包含了这种功能, 广义线性模型在一般的工业研究与开发领域中会有更加宽广的应用.

15.2 析因设计中的不平衡数据

本章的基本目的是分析**平衡析因设计**, 即在每个单元处观测值的个数都是相同的. 其实, 观测值个数不相等的情况也经常遇到. 导致不平衡**析因设计**的原因很多. 例如, 实验者起初设计了一个平衡实验, 但因为在收集数据时出现不可预见的问题, 以致于丢失了一些观测值, 她或他最后得到的就是不平衡数据. 另一方面, 有些不平衡实验是特意设计的. 比方说, 某些处理组合比起其他的组合做起试验来更为昂贵或更为困难, 这样, 在这些单元中就会取较少的观测值. 或者, 实验者对某些处理组合有较大的兴趣, 因为它们代表了新的或未研究过的条件, 所以, 实验者会决定在这些单元中多做几次重复.

在平衡数据中出现的主效应以及交互作用的正交性质在不非平衡情况下将不再具备. 这意味着我们不能再使用通常的方差分析法. 因此, 不平衡析因设计比起平衡设计的分析要困难得多.

本节简要综述一下关于处理不平衡析因设计的方法, 着重于二因子固定效应模型的情况. 设在第 ij 单元的观测值个数为 n_{ij}. 又令 $n_{i.}=\sum\limits_{j=1}^{b}n_{ij}$ 为第 i 行 (因子 A 的第 i 个水平) 中观测

值的个数, $n_{.j}=\sum_{i=1}^{a}n_{ij}$ 为第 j 列 (因子 B 的第 j 个水平) 中观测值的个数, 而 $n_{..}=\sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij}$ 是观测值的总个数.

15.2.1　成比例的数据: 简单的情况

在分析中难度不大的一种不平衡数据情况是**成比例的数据**. 也就是说, 第 ij 单元中的观测值的个数是

$$n_{ij}=\frac{n_{i.}n_{.j}}{n_{..}} \tag{15.11}$$

这一条件意味着任意两行或两列的观测值的个数是成比例的. 当出现比例数据时, 可以用标准的方差分析. 只须在平方和的计算公式中作小的修改即可, 它们是

$$SS_T=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}y_{ijk}^2-\frac{y_{...}^2}{n_{..}},\qquad SS_{AB}=\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{y_{ij.}^2}{n_{ij}}-\frac{y_{...}^2}{n_{..}}-SS_A-SS_B$$

$$SS_A=\sum_{i=1}^{a}\frac{y_{i..}^2}{n_{i.}}-\frac{y_{...}^2}{n_{..}},\qquad SS_E=SS_T-SS_A-SS_B-SS_{AB}$$

$$SS_B=\sum_{j=1}^{b}\frac{y_{.j.}^2}{n_{.j}}-\frac{y_{...}^2}{n_{..}},\qquad =\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}y_{ijk}^2-\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{y_{ij.}^2}{n_{ij}}$$

这样就得到一个基于序贯模型拟合分析的 ANOVA, 先拟合因子 A 再拟合因子 B(另一种方法是采用 "调整的" 模型拟合方法, 类似于第 4 章对不完全区组设计所用的方法: 两种方法都可以用 Minitab Balanced ANOVA 程序来实现).

作为比例数据的例子, 考虑例 5.1 的电池设计实验. 表 15.6 是原始数据的一种修正. 显然, 这些数据是成比例的. 例如, 在单元 1,1 中有

$$n_{11}=\frac{n_{1.}n_{.1}}{n_{..}}=\frac{10(8)}{20}=4$$

个观测值. 对这些数据应用通常的方差分析, 其结果如表 15.7 所示. 材料类型和温度都是显著的, 而交互作用仅在 $\alpha=0.17$ 时才是显著的. 因此, 除了交互作用的效应不显著外, 结论与例 5.1 中对完全数据集的分析结果大体一致.

表 15.6　比例数据的电池设计实验

材料类型	温度/°F 15		70		125		
1	$n_{11}=4$		$n_{12}=4$		$n_{13}=2$		$n_{1.}=10$
	130	155	34	40			
	74	180	80	75	70	58	$y_{1..}=896$
2	$n_{21}=2$		$n_{22}=2$		$n_{23}=1$		$n_{2.}=5$
	159	126	136	115	45		$y_{2..}=581$
3	$n_{31}=2$		$n_{32}=2$		$n_{33}=1$		$n_{3..}=5$
	138	160	150	139	96		$y_{3..}=683$
	$n_{.1}=8$		$n_{.2}=8$		$n_{.3}=4$		$n_{..}=20$
	$y_{.1.}=1\ 122$		$y_{.2.}=769$		$y_{.3.}=269$		$y_{...}=2\ 160$

表 15.7 表 15.6 中电池设计数据的方差分析

方差来源	平方和	自由度	均方	F_0
材料类型	7 811.6	2	3 905.8	4.78
温度	16 090.9	2	8 045.5	9.85
交互作用	6 266.5	4	1 566.6	1.92
误差	8 981.0	11	816.5	
总和	39 150.0	19		

15.2.2 近似方法

当不平衡数据较接近于平衡情况时, 有时可以用**近似方法**把不平衡问题转变为平衡问题. 当然, 这样只能作出近似的分析, 但因为平衡数据的分析很容易, 所以我们经常试图采用这一方法. 在实践上, 当数据与平衡情况不是太不一样时, 确定所引入方法的近似程度相对来说就不重要了. 现在简要描述一些近似方法. 假定每一单元至少有一个观测值 (即 $n_{ij} \geqslant 1$).

1. 估计缺失观测值

如果只有少数 n_{ij} 不相同, 合理的方法是估计缺失值. 例如, 考虑表 15.8 的非平衡设计. 显然, 估计单元 2,2 的单个缺失值是一种合理的方法. 对于有交互作用的模型, 能够最小化误差平方和的第 ij 单元的缺失值的估计是 $\bar{y}_{ij.}$. 也就是说, 用那一单元中现有的观测值的平均值来估计缺失值.

表 15.8 非平衡设计的 n_{ij} 值

行	列		
	1	2	3
1	4	4	4
2	4	3	4
3	4	4	4

将估计值作为实际数据处理. 方差分析时唯一要修改的是, 将误差自由度减去已估计的缺失观测值的个数. 例如, 如果估计出表 15.8 中 2,2 单元的缺失值, 则用 26 个误差自由度来代替 27 个误差自由度.

2. 搁置数据

考虑表 15.9 中的数据. 只有 2,2 单元比其他单元多一个观测值. 此时, 估计其余 8 个缺失值就不是一个好主意, 因为这导致估计值在最后的数据中占了约 18%. 另一种方法是把单元 2,2 的一个观测值搁置一旁, 而得出一个有 $n = 4$ 次重复的平衡设计.

表 15.9 非平衡设计的 n_{ij} 值

行	列		
	1	2	3
1	4	4	4
2	4	5	4
3	4	4	4

搁置一旁的观测值必须是随机选取的, 不但不能完全丢弃它, 我们还将把它重新放回到设计中, 然后再随机选取另一观测值搁置一旁, 再进行分析. 并且, 我们希望, 这两次分析不会得出关于数据的互相矛盾的解释. 如果出现矛盾, 则我们怀疑已被搁置一旁的观测值是一个异常值或野值, 因而应当做出相应处理. 在实践中, 当仅有少数的观测值被搁置一旁, 而且单元内部的变异性较小时, 这种混乱情况不大可能出现.

3. 未加权平均数法

在 Yates (1934) 所引入的这一近似法中, 把单元平均值看作为数据并按标准的平衡数据分析方法来求得行、列以及交互作用的平方和. 求得误差均方为

$$MS_E = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(y_{ijk} - \bar{y}_{ij.})^2}{n_{..} - ab} \tag{15.12}$$

现在, 用 MS_E 估计 σ^2, 这是单个观测值 y_{ijk} 的方差. 然而我们所做的是关于单元平均值的方差分析. 因为第 ij 单元的平均值的方差是 σ^2/n_{ij}, 所以实际用在方差分析中的误差均方值应该是 $\bar{y}_{ij.}$ 的平均方差的一个估计量, 即

$$\bar{\mathrm{V}}(\bar{y}_{ij.}) = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sigma^2/n_{ij}}{ab} = \frac{\sigma^2}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}} \tag{15.13}$$

用 (15.12) 式的 MS_E 来估计 (15.13) 式的 σ^2, 得

$$MS_E' = \frac{MS_E}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij}} \tag{15.14}$$

作为用在方差分析中的误差均方值 (有 $n_{..} - ab$ 个自由度).

未加权平均数法是一种近似方法, 因为行、列以及交互作用的平方和都不服从卡方分布. 这一方法的主要优点是它在计算上的简明性. 当 n_{ij} 不是相差很多时, 未加权平均数法的效果通常不错.

一种相关的方法是**加权平均数法**, 也是 Yates (1934) 提出的. 这一方法也用到单元平均数的平方和, 但是, 平方和中的项是反比于相应方差进行加权的. 这一方法更详细的叙述, 可参阅 Searle(1971a) 以及 speed, Hocking, and Hackney(1978).

15.2.3　精确方法

当近似方法不适用时, 例如, 当出现空的单元 (某些 $n_{ij} = 0$) 时, 或者当 n_{ij} 有很大差别时, 实验者必须用精确分析法. 为求出用来检验主效应与交互作用效应的平方和, 做法是把方差分析模型作为**回归模型**, 对数据拟合该模型, 并用一般回归显著性检验方法. 然而, 有多种方式可做到这一点, 而且这些方法会得出不同的平方和数值. 再说, 需要检验的假设也不总是类似于平衡情况, 且它们也并不总是易于解释的. 要进一步阅读这一主题, 可参阅 Searle(197la), Speed, and Hocking(1976), Hocking and Speed(1975), Hocking, Hackney, and Speed(1978), Searle, Speed, and Henderson(1981), Searle(1987), 以及 Milliken and Johnson(1984). 统计软件 SAS 通过 PROC GLM 为不平衡数据分析提供了极好的办法.

15.3 协方差分析

第 2 章与第 4 章介绍了区组化原则的使用, 以提高处理间比较的精度. 配对 t 检验是第 2 章中介绍的方法, 随机化区组设计是第 4 章给出的. 一般来讲, **区组化原则**能用于可控讨厌因子的影响. **协方差分析**(ANCOVA) 是另一种常用于改进实验精度的技术. 设在一实验中, 响应变量是 y, 存在另一变量, 例如 x, 且 y 线性依赖于 x. 还有, 设 x 不能被实验者控制, 但可以随着 y 一起被观测到. 变量 x 叫做**协变量或伴随变量**. 协方差分析涉及如何针对协变量的效应调节被观测的响应变量. 如果不做这样的调节, 则伴随变量会提高误差均方值, 并使由于处理不同而引起的响应之间的真正差异更难于检测. 这样一来, 协方差分析就是调节不可控讨厌变量效应的一种方法. 正如我们将会见到的, 这一方法是方差分析和回归分析的综合.

作为可以使用协方差分析的一个实验例子, 考虑进行一项研究, 以确定由 3 台不同机器生产的单丝纤维的强度的差异. 该实验数据列在表 15.10 中. 图 15.3 是样品的抗断强度与直径 (粗细) 关系的散点图. 显然, 纤维的抗断强度亦受其纤度或粗细所影响, 较粗的纤维一般强于较细的纤维. 当检验机器之间的抗断强度的差异时, 协方差分析可以用来消除粗细度 (x) 对抗断强度 (y) 的效应.

表 15.10 抗断强度数据: $y=$ 强度 (磅), $x=$ 直径 (10^{-3} 英寸)

机器 1		机器 2		机器 3	
y	x	y	x	y	x
36	20	40	22	35	21
41	25	48	28	37	23
39	24	39	22	42	26
42	25	45	30	34	21
49	32	44	28	32	15
207	126	216	130	180	106

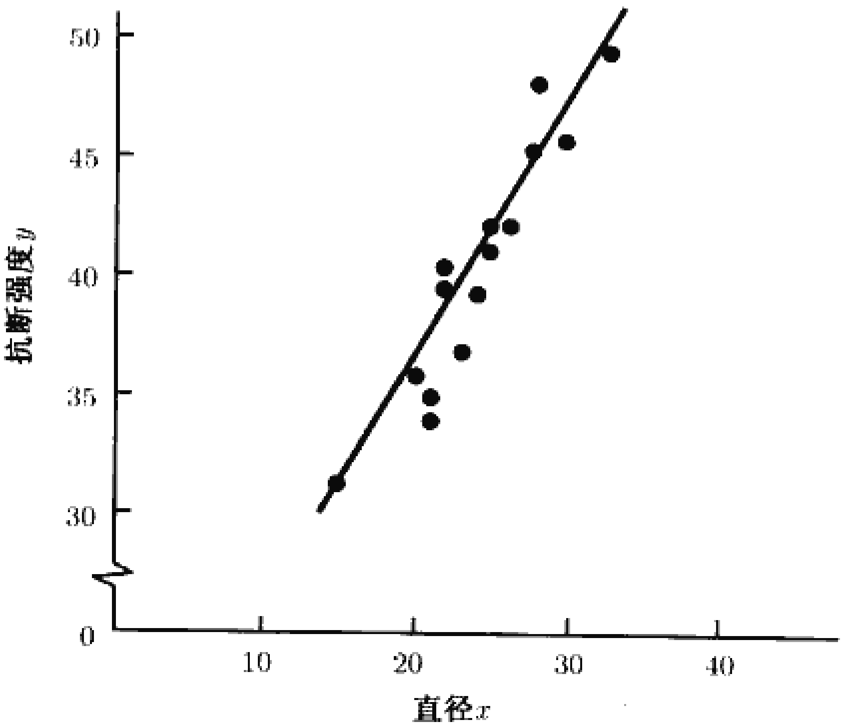


图 15.3 抗断强度 (y) 与纤维直径 (x) 的关系

15.3.1 方法说明

现在针对有一个协变量的单因子实验来说明协方差分析的基本方法. 假定响应和协变量之间存在线性关系, 恰当的统计模型是

$$y_{ij}=\mu+\tau_i+\beta(x_{ij}-\bar{x}_{..})+\varepsilon_{ij}\qquad\begin{cases}i=1,2,\cdots,a\\ j=1,2,\cdots,n\end{cases}\tag{15.15}$$

其中 y_{ij} 是在单因子的第 i 种处理或水平下取得的响应变量的第 j 个观测值, x_{ij} 是对应于 y_{ij} (即, 第 ij 个试验) 的协变量或伴随变量的测量值, $\bar{x}_{..}$ 是全体 x_{ij} 的平均值, μ 是总平均值, τ_i 是第 i 种处理的效应, β 是线性回归系数, 表示 y_{ij} 对 x_{ij} 的相依性, ε_{ij} 是随机误差分量. 假定误差 ε_{ij} 是 NID$(0,\sigma^2)$ 的, 斜率 $\beta\neq0$, y_{ij} 与 x_{ij} 之间的真实关系是线性的, 每种处理的回归系数是相同的, 处理效应之和为零 $\left(\sum\limits_{i=1}^{a}\tau_i=0\right)$, 伴随变量 x_{ij} 不受处理的影响.

由 (15.15) 式看出, 协方差分析模型是方差分析和回归分析所用的线性模型的组合. 也就是说, 有单因子方差分析的处理效应 $\{\tau_i\}$ 和回归分析的回归系数 β. (15.15) 式中用 $(x_{ij}-\bar{x}_{..})$ 代替 x_{ij} 来表示伴随变量, 这样可使参数 μ 是总平均值. 模型原先可写为

$$y_{ij}=\mu'+\tau_i+\beta x_{ij}+\varepsilon_{ij}\qquad\begin{cases}i=1,2,\cdots,a\\ j=1,2,\cdots,n\end{cases}\tag{15.16}$$

其中 μ' 是常数, 不等于总平均值, 此模型的总平均值 $\mu'+\beta\bar{x}_{..}$. 在文献中更多见到的是 (15.15) 式.

为了描述协方差分析, 引入下列记号:

$$S_{yy}=\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2=\sum_{i=1}^{a}\sum_{j=1}^{n}y_{ij}^2-\frac{y_{..}^2}{an}\tag{15.17}$$

$$S_{xx}=\sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{..})^2=\sum_{i=1}^{a}\sum_{j=1}^{n}x_{ij}^2-\frac{x_{..}^2}{an}\tag{15.18}$$

$$S_{xy}=\sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{..})(y_{ij}-\bar{y}_{..})=\sum_{i=1}^{a}\sum_{j=1}^{n}x_{ij}y_{ij}-\frac{(x_{..})(y_{..})}{an}\tag{15.19}$$

$$T_{yy}=n\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2=\frac{1}{n}\sum_{i=1}^{a}y_{i.}^2-\frac{y_{..}^2}{an}\tag{15.20}$$

$$T_{xx}=n\sum_{i=1}^{a}(\bar{x}_{i.}-\bar{x}_{..})^2=\frac{1}{n}\sum_{i=1}^{a}x_{i.}^2-\frac{x_{..}^2}{an}\tag{15.21}$$

$$T_{xy}=n\sum_{i=1}^{a}(\bar{x}_{i.}-\bar{x}_{..})(\bar{y}_{i.}-\bar{y}_{..})=\frac{1}{n}\sum_{i=1}^{a}(x_{i.})(y_{i.})-\frac{(x_{..})(y_{..})}{an}\tag{15.22}$$

$$E_{yy}=\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2=S_{yy}-T_{yy}\tag{15.23}$$

$$E_{xx}=\sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{i.})^2=S_{xx}-T_{xx}\tag{15.24}$$

$$E_{xy}=\sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{i.})(y_{ij}-\bar{y}_{i.})=S_{xy}-T_{xy}\tag{15.25}$$

注意, 一般地, $S = T + E$, 其中符号 S, T, E 分别用于表示总、处理以及误差的平方和与叉积和. 对 x 和 y 的平方和必须是非负的; 但是, 叉积 (xy) 的和可以是负的.

以下说明, 协方差分析怎样根据协变量效应来调整响应变量. 考虑全模型 [(15.15) 式]. μ, τ_i, β 的最小二乘估计是 $\hat{\mu} = \bar{y}_{..}$, $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(x_{i.} - \bar{x}_{..})$ 以及

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}} \tag{15.26}$$

模型中的误差平方和为

$$SS_E = E_{yy} - (E_{xy})^2/E_{xx} \tag{15.27}$$

它有 $a(n-1)-1$ 个自由度. 实验的误差方差用

$$MS_E = \frac{SS_E}{a(n-1)-1}$$

来估计. 今假设不存在处理效应. 则模型 [(15.15) 式] 将是

$$y_{ij} = \mu + \beta(x_{ij} - \bar{x}_{..}) + \epsilon_{ij} \quad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, n \end{cases} \tag{15.28}$$

μ 与 β 的最小二乘估计是 $\hat{\mu} = \bar{y}_{..}$, $\hat{\beta} = S_{xy}/S_{xx}$. 此简化模型的误差平方和是

$$SS'_E = S_{yy} - (S_{xy})^2/S_{xx} \tag{15.29}$$

它有 $an-2$ 个自由度. 在 (15.29) 式中, $(S_{xy})^2/S_{xx}$ 是通过 y 关于 x 的线性回归所获得的、在 y 的平方和中的减少量. 还有, 注意到 SS_E 小于 SS'_E[因为模型 (15.15) 式含有附加的参数 $\{\tau_i\}$]. 量 $SS'_E - SS_E$ 是由于 $\{\tau_i\}$ 而在平方和中的减少量. 因此 SS'_E 与 SS_E 之间的差, 即 $SS'_E - SS_E$ 为检验没有处理效应的假设提供了有 $a-1$ 个自由度的平方和. 因此, 为了检验 $H_0: \tau_i = 0$, 计算

$$F_0 = \frac{(SS'_E - SS_E)/(a-1)}{SS_E/[a(n-1)-1]} \tag{15.30}$$

如果零假设为真, 则它服从 $F_{a-1,a(n-1)-1}$ 分布. 当 $F_0 > F_{\alpha,a-1,a(n-1)-1}$ 时, 拒绝 $H_0: \tau_i = 0$. 也可以用 P 值方法.

考察表 15.11 的内容是有益的. 在此表中, 我们已介绍过将协方差分析作为 "调整" 的方差分析. 在 "方差来源" 一列, 总变异性用 S_{yy} 来测量, 它有 $an-1$ 个自由度. 方差来源 "回归" 是有一个自由度的平方和 $(S_{xy})^2/S_{xx}$. 如果没有协变量, 将会有 $S_{xy} = S_{xx} = E_{xy} = E_{xx} = 0$.

表 15.11 作为 "调整" 的方差分析的协方差分析

方差来源	平方和	自由度	均方	F_0
回归	$(S_{xy})^2/S_{xx}$	1		
处理	$SS'_E - SS_E = S_{yy} - (S_{xy})^2/S_{xx} - [E_{yy} - (E_{xy})^2/E_{xx}]$	$a-1$	$\frac{SS'_E - SS_E}{a-1}$	$\frac{(SS'_E - SS_E)/(a-1)}{MS_E}$
误差	$SS_E = E_{yy} - (E_{xy})^2/E_{xx}$	$a(n-1)-1$	$MS_E = \frac{SS_E}{a(n-1)-1}$	
总和	S_{yy}	$an-1$		

误差平方和将简化为 E_{yy}, 而处理平方和将是 $S_{yy} - E_{yy} = T_{yy}$. 但是, 因为出现协变量, 就必须如表 15.11 那样对 y 在 x 上的回归 "调节" S_{yy} 和 E_{yy}. 已调节的误差平方和有 $a(n-1)-1$ 个自由度而不是 $a(n-1)$ 个自由度, 因为有一个附加的参数 (斜率 β) 用于拟合这些数据.

计算方法通常以表 15.12 那样的协方差分析表来显示. 采用这一布局是因为它便于概括出所有要求的平方和与叉积和, 以及为检验关于处理效应的假设的平方和. 此外, 除了检验处理效应没有差异的假设之外, 经常发现, 给出调整的处理均值对解释数据是有用的. 这些已调整的处理均值是按

$$\text{调整的 } \bar{y}_{i.} = \bar{y}_{i.} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \quad i = 1, 2, \cdots, a \tag{15.31}$$

计算的, 其中 $\hat{\beta} = E_{xy}/E_{xx}$. 这一已调整的处理平均值就是模型 [(15.15) 式] 中的 $\mu + \tau_i$, $i = 1, 2, \cdots, a$ 的最小二乘估计量. 任一已调整的处理平均值的标准误是

$$S_{\text{调整的}\bar{y}_{i.}} = \left[MS_E\left(\frac{1}{n} + \frac{(\bar{x}_{i.} - \bar{x}_{..})^2}{E_{xx}}\right)\right]^{1/2} \tag{15.32}$$

最后, 我们曾经假定模型 [(15.15) 式] 的回归系数 β 是非零的. 可以用检验统计量

$$F_0 = \frac{(E_{xy})^2/E_{xx}}{MS_E} \tag{15.33}$$

来检验假设 $H_0: \beta = 0$, 在零假设下, 它的分布为 $F_{1,a(n-1)-1}$ 分布. 因此, 当 $F_0 > F_{\alpha,1,a(n-1)-1}$ 时, 拒绝 $H_0: \beta = 0$.

表 15.12　有一个协变量的单因子实验的协方差分析

方差来源	自由度	平方和与叉积和			已对回归调整的		
		x	xy	y	y	自由度	均方
处理	$a-1$	T_{xx}	T_{xy}	T_{yy}			
误差	$a(n-1)$	E_{xx}	E_{xy}	E_{yy}	$SS_E = E_{yy} - (E_{xy})^2/E_{xx}$	$a(n-1)-1$	$MS_E = \frac{SS_E}{a(n-1)-1}$
总和	$an-1$	S_{xx}	S_{xy}	S_{yy}	$SS'_E = S_{yy} - (S_{xy})^2/S_{xx}$	$an-2$	
调整的处理均值					$SS'_E - SS_E$	$a-1$	$\frac{SS'_E - SS_E}{a-1}$

例 15.5　考虑 15.3 节开始时描述的实验. 3 台不同的机器为纺织公司生产单丝纤维. 过程工程师感兴趣于确定 3 台机器生产的纤维的抗断强度是否有差异. 但是, 纤维的强度与它的直径有关, 较粗的纤维一般强于较细的. 从每台机器上选取 5 根纤维样品的随机样本. 每一样品的纤维强度 (y) 和对应的直径 (x) 如表 15.10 所示.

抗断强度与纤维直径关系的散点图 (图 15.3) 强烈提示我们, 抗断强度与直径之间存在线性关系, 而用协方差分析消除直径对强度的效应看起来是合适的. 假定抗断强度与直径之间是一线性关系是恰当的, 则模型是

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, 3 \\ j = 1, 2, \cdots, 5 \end{cases}$$

利用 (15.17) 式至 (15.25) 式, 可算得

$$S_{yy} = \sum_{i=1}^{3}\sum_{j=1}^{5} y_{ij}^2 - \frac{y_{..}^2}{an} = (36)^2 + (41)^2 + \cdots + (32)^2 - \frac{(603)^2}{(3)(5)} = 346.40$$

$$S_{xx} = \sum_{i=1}^{3}\sum_{j=1}^{5} x_{ij}^2 - \frac{x_{..}^2}{an} = (20)^2 + (25)^2 + \cdots + (15)^2 - \frac{(362)^2}{(3)(5)} = 261.73$$

$$S_{xy} = \sum_{i=1}^{3}\sum_{j=1}^{5} x_{ij}y_{ij} - \frac{(x_{..})(y_{..})}{an} = (20)(36) + (25)(41) + \cdots + (15)(32)$$

$$-\frac{(362)(603)}{(3)(5)} = 282.60$$

$$T_{yy} = \frac{1}{n}\sum_{i=1}^{3} y_{i.}^2 - \frac{y_{..}^2}{an} = \frac{1}{5}[(207)^2 + (216)^2 + (180)^2] - \frac{(603)^2}{(3)(5)} = 140.40$$

$$T_{xx} = \frac{1}{n}\sum_{i=1}^{3} x_{i.}^2 - \frac{x_{..}^2}{an} = \frac{1}{5}[(126)^2 + (130)^2 + (106)^2] - \frac{(362)^2}{(3)(5)} = 66.13$$

$$T_{xy} = \frac{1}{n}\sum_{i=1}^{3} x_{i.}y_{i.} - \frac{(x_{..})(y_{..})}{an}$$

$$= \frac{1}{5}[(126)(207) + (130)(216) + (106)(184)] - \frac{(362)(603)}{(3)(5)} = 96.00$$

$$E_{yy} = S_{yy} - T_{yy} = 346.40 - 140.40 = 206.00$$

$$E_{xx} = S_{xx} - T_{xx} = 261.73 - 66.13 = 195.60$$

$$E_{xy} = S_{xy} - T_{xy} = 282.60 - 96.00 = 186.60$$

由 (15.29) 式, 得

$$SS_E' = S_{yy} - (S_{xy})^2/S_{xx} = 346.40 - (282.60)^2/261.73 = 41.27$$

它有 $an - 2 = (3)(5) - 2 = 13$ 个自由度. 由 (15.27) 式, 求得

$$SS_E = E_{yy} - (E_{xy})^2/E_{xx} = 206.00 - (186.60)^2/195.60 = 27.99$$

它有 $a(n-1) - 1 = 3(5-1) - 1 = 11$ 个自由度.

用于检验 $H_0: \tau_1 = \tau_2 = \tau_3 = 0$ 的平方和是

$$SS_E' - SS_E = 41.27 - 27.99 = 13.28$$

它有 $a - 1 = 3 - 1 = 2$ 个自由度. 这些计算列在表 15.13 中.

表 15.13 抗断强度数据的方差分析

方差来源	自由度	平方和与叉积和			对回归的调节			F_0	P 值
		x	y	xy	y	自由度	均方		
机器	2	66.13	96.00	140.40					
误差	12	195.60	186.60	206.00	27.99	11	2.54		
总和	14	261.73	282.60	346.40	41.27	13			
已调节的机器					13.28	2	6.64	2.61	0.1181

为了检验机器生产的纤维的抗断强度的差异性的假设 (即, $H_0: \tau_i = 0$), 由 (15.30) 式算得的检验统计量为

$$F_0 = \frac{(SS_E' - SS_E)/(a-1)}{SS_E/[a(n-1)-1]} = \frac{13.28/2}{27.99/11} = \frac{6.64}{2.54} = 2.61$$

与 $F_{0.10,2,11} = 2.86$ 比较, 我们发现不能拒绝零假设. 也就是说, 没有强有力的证据表明 3 台机器所生产的纤维的抗断强度有差异.

由 (15.26) 式算得回归系数的估计量为

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}} = \frac{186.60}{195.60} = 0.954\ 0$$

用 (15.33) 式检验假设 $H_0 : \beta = 0$. 检验统计量是

$$F_0 = \frac{(E_{xy})^2/E_{xx}}{MS_E} = \frac{(186.60)^2/195.60}{2.54} = 70.08$$

因为 $F_{0.01,1,11} = 9.65$, 我们拒绝 $\beta = 0$ 的假设. 因此, 抗断强度和直径之间存在线性关系, 用协方差分析进行的调整是必需的.

用 (15.31) 式计算调整的处理平均值. 这些调整的均值是

调整的 $\bar{y}_{1.} = \bar{y}_{1.} - \hat{\beta}(\bar{x}_{1.} - \bar{x}_{..}) = 41.40 - (0.954\ 0)(25.20 - 24.13) = 40.38$

调整的 $\bar{y}_{2.} = \bar{y}_{2.} - \hat{\beta}(\bar{x}_{2.} - \bar{x}_{..}) = 43.20 - (0.954\ 0)(26.00 - 24.13) = 41.42$

调整的 $\bar{y}_{3.} = \bar{y}_{3.} - \hat{\beta}(\bar{x}_{3.} - \bar{x}_{..}) = 36.00 - (0.954\ 0)(21.20 - 24.13) = 38.80$

将已调整的处理平均值与未调整的处理平均值 ($\bar{y}_{i.}$) 进行比较, 看出已调整的平均值相互较为靠近, 这从另一角度表明协方差分析是必需的.

协方差分析的基本假定是处理不影响协变量 x, 因为此方法消除了 $\bar{x}_{i.}$ 的变异效应. 但是, 若 $\bar{x}_{i.}$ 的变异性部分来自于处理, 则协方差分析就丢弃了处理效应的一部分. 这样一来, 我们必须相当确信处理不影响数值 x_{ij}. 在一些实验中, 由协变量的性质, 这点可能是显然的, 而在另一些实验中, 可能拿不准. 在我们的例子中, 3 台机器之间的纤维直径 (x_{ij}) 可能存在差异. 在这种情况下. Cochran and Cox(1957) 提出了一种关于数值 x_{ij} 的方差分析法, 它有助于确定这个假定的有效性. 对于我们的问题, 此方法得出

$$F_0 = \frac{66.13/2}{195.60/12} = \frac{33.07}{16.30} = 2.03$$

它小于 $F_{0.10,2,12} = 2.81$, 所以没有理由相信那些机器生产出不同直径的纤维.

协方差模型的诊断检测是建立在残差分析的基础上的. 对于协方差模型, 残差是

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$

其中的拟合值是

$$\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}(x_{ij} - \bar{x}_{..}) = \bar{y}_{..} + [\bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})] + \hat{\beta}(x_{ij} - \bar{x}_{..})$$
$$= \bar{y}_{i.} + \hat{\beta}(x_{ij} - \bar{x}_{i.})$$

于是,

$$e_{ij} = y_{ij} - \bar{y}_{i.} - \hat{\beta}(x_{ij} - \bar{x}_{i.}) \tag{15.34}$$

为说明 (15.34) 式的用法, 在例 15.5 中, 第一台机器的第一个观测值的残差是

$$e_{11} = y_{11} - \bar{y}_{1.} - \hat{\beta}(x_{11} - \bar{x}_{1.}) = 36 - 41.4 - (0.954\ 0)(20 - 25.2) = 36 - 36.439\ 2 = -0.439\ 2$$

全部观测值、拟合值和残差列于下表:

观测值 y_{ij}	拟合值 $\hat{y}_{ij}$	残差 $e_{ij} = y_{ij} - \hat{y}_{ij}$	观测值 y_{ij}	拟合值 $\hat{y}_{ij}$	残差 $e_{ij} = y_{ij} - \hat{y}_{ij}$
36	36.439 2	−0.439 2	42	41.209 2	0.790 8
41	41.209 2	−0.209 2	49	47.887 1	1.112 9
39	40.255 2	−1.255 2	40	39.384 0	0.616 0

(续)

观测值 y_{ij}	拟合值 $\hat{y}_{ij}$	残差 $e_{ij}=y_{ij}-\hat{y}_{ij}$	观测值 y_{ij}	拟合值 $\hat{y}_{ij}$	残差 $e_{ij}=y_{ij}-\hat{y}_{ij}$
48	45.107 9	2.982 1	37	37.717 1	−0.717 1
39	39.384 0	−0.384 0	42	40.579 1	1.420 9
45	47.015 9	−2.015 9	34	35.809 2	−1.809 2
44	45.107 9	−1.107 9	32	30.085 2	1.914 8
35	35.809 2	−0.809 2			

拟合值 $\hat{y}_{ij}$ 与残差的关系图见图 15.4, 协变量 x_{ij} 与残差的关系图见图 15.5, 机器与残差的关系图见图 15.6. 残差的正态概率图见图 15.7. 这些图形没有显现出与假定有任何违背之处, 所以, 结论是, 协方差模型 [(15.15) 式] 对于抗断强度数据是适当的.

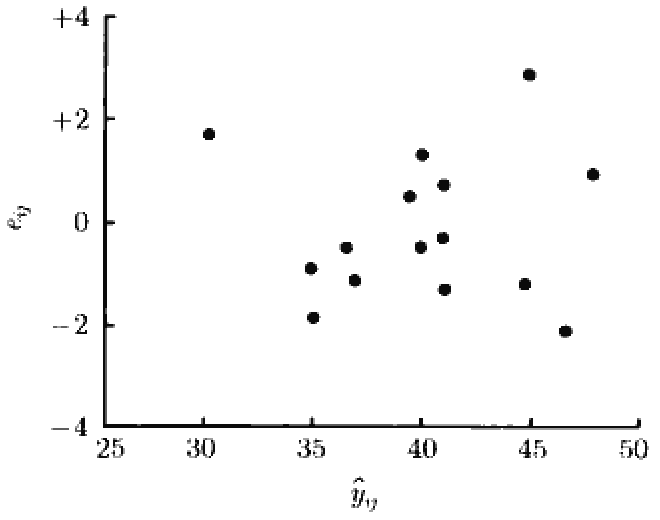


图 15.4 例 15.5 的残差与拟合值的关系图

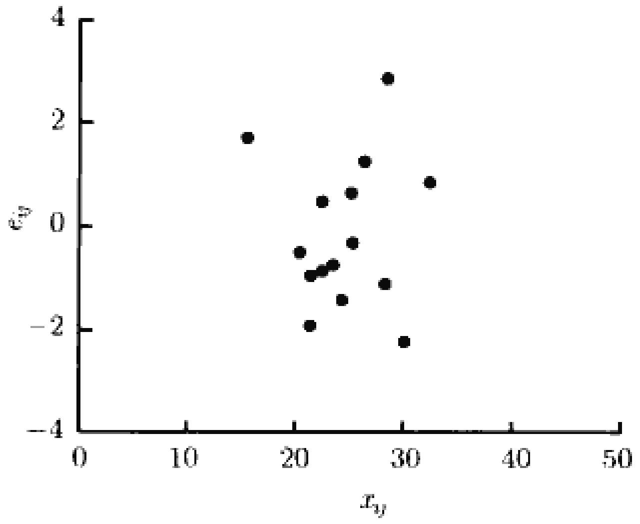


图 15.5 例 15.5 的残差与纤维直径 x 的关系图

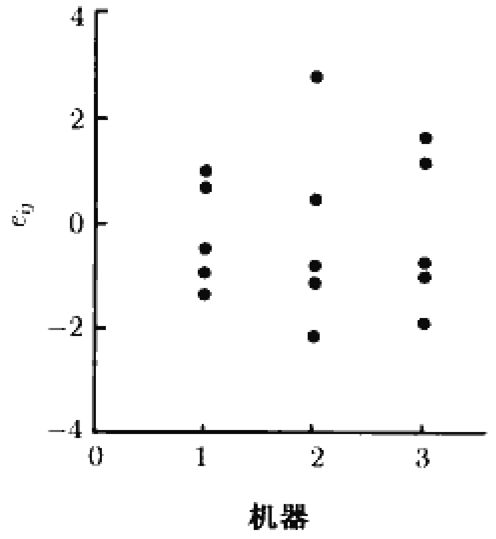


图 15.6 残差与机器的关系图

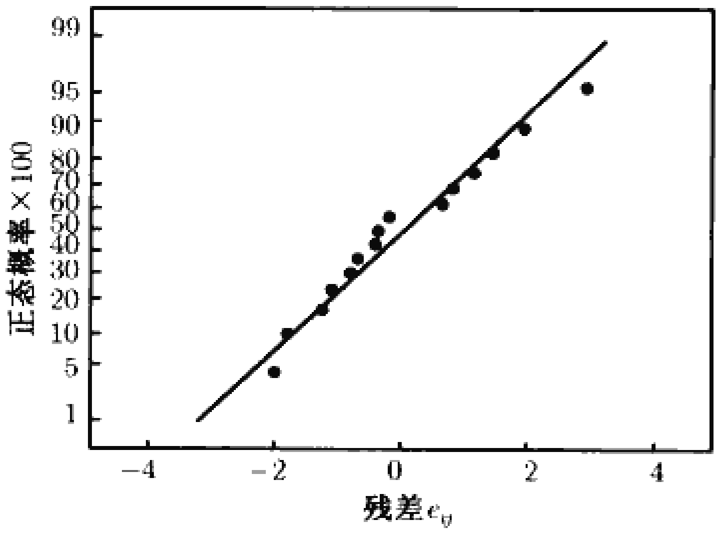


图 15.7 例 15.5 的残差的正态概率图

在这一实验中, 如果没有进行协方差分析, 那么会什么样呢? 也就是说, 如果略去协变量 x, 将抗断强度数据 (y) 作为单因子实验来分析, 会发生什么样的情况? 抗断强度数据的方差分析如表 15.14 所示. 据此分析, 结论是机器生产的纤维的强度有显著性差异. 这与由协方差分析所得的结论完全相反. 如果推测机器对纤维强度的效应有显著性差异, 则我们应该试图使 3 台机器的强度输出相等. 但是, 在这一问题中, 在消除了纤维直径的线性效应之后, 几台机器所生产的纤维的强度是没有差异的. 而降低机器内部纤维直径的变异性可能会有用, 因为这有可能降低纤

维强度的变异性.

表 15.14　抗断强度数据作为单因子实验的不正确分析

方差来源	平方和	自由度	均方	F_0	P 值
机器	140.40	2	70.20	4.09	0.044 2
误差	206.00	12	17.17		
总和	346.40	14			

15.3.2　计算机输出

现有的几个计算机软件包能够进行协方差分析. 关于例 15.4 的数据的 Minitab General Linear Models 程序输出列在表 15.15 中. 该输出与前面给出的结果非常相似. 在题为 "Analysis of Variance" 的部分中, "Seq SS" 对应于模型总平方和的"序贯"分解, 即

$$SS(\text{模型}) = SS(\text{直径}) + SS(\text{机器}|\text{直径}) = 305.13 + 13.28 = 318.41$$

其中 "Adj SS" 对应于每个因子的 "额外" 平方和, 即

$$SS(\text{机器}|\text{直径}) = 13.28, \qquad SS(\text{直径}|\text{机器}) = 178.01$$

表 15.15　例 15.5 的 Minitab 输出 (协方差分析)

```
General Linear Model

Factor     Type    Levels   Values
Machine    fixed        3   1  2  3

Analysis of Variance for Strength, using Adjusted SS for Tests

Source     DF     Seq SS      Adj SS      Adj MS        F        P
Diameter    1     305.13      178.01      178.01    69.97    0.000
Machine     2      13.28       13.28        6.64     2.61    0.118
Error      11      27.99       27.99        2.54
Total      14     346.40

Term          Coef         StDev        T          P
Constant     17.177        2.783      6.17      0.000
Diameter     0.9540       0.1140      8.36      0.000
Machine
1            0.1824       0.5950      0.31      0.765
2            1.2192       0.6201      1.97      0.075

Means  for  Covariates

Covariate      Mean       StDey
Diameter      24.13       4.324

Least Squares Means for Strength

Machine      Mean     StDev
1           40.38    0.7236
2           41.42    0.7444
3           38.80    0.7879
```

注意, SS(机器|直径) 用于检验没有机器效应的校正平方和, SS(机器|直径) 用于检验假设 $\beta = 0$ 的校正平方和. 因为取整, 表 15.15 中的检验统计量与那些手工算的结果稍有不同.

程序也计算 (15.31) 式的调整的处理均值 [样例输出中 Minitab 称之为最小平方均值 (least squares means)] 以及标准误. 程序还会用第 3 章讨论的配对多重比较方法对所有处理配对进行比较.

15.3.3 用一般线性回归显著性检验进行推导

在协方差模型

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, n \end{cases} \tag{15.35}$$

中, 利用一般回归显著性检验法, 可以正式推导出用于检验 $H_0 : \tau_i = 0$ 的 ANCOVA 方法. 考虑用最小二乘法估计模型 [(15.15) 式] 中的参数. 最小二乘函数为

$$L = \sum_{i=1}^{a} \sum_{j=1}^{n} [y_{ij} - \mu - \tau_i - \beta(x_{ij} - \bar{x}_{..})]^2 \tag{15.36}$$

由 $\partial L/\partial\mu = \partial L/\partial\tau_i = \partial L/\partial\beta = 0$, 得到正规方程组

$$\mu : an\hat{\mu} + n\sum_{i=1}^{a} \hat{\tau}_i = y_{..} \tag{15.37a}$$

$$\tau_i : n\hat{\mu} + n\hat{\tau}_i + \hat{\beta}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = y_{i.} \quad i = 1, 2, \cdots, a \tag{15.37b}$$

$$\beta : \sum_{i=1}^{a} \hat{\tau}_i \sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) + \hat{\beta}S_{xx} = S_{xy} \tag{15.37c}$$

把 (15.37b) 中的 a 个方程相加, 因为 $\sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = 0$, 故得到 (15.37a) 式. 所以, 在正规方程组中有一个线性相依性. 因此, 为了求解, 就必须给 (15.37) 式添加一个线性独立的方程. 一个合乎逻辑的边界条件是 $\sum_{i=1}^{a} \hat{\tau}_i = 0$.

用此条件, 由 (15.37a) 式求得

$$\hat{\mu} = \bar{y}_{..} \tag{15.38a}$$

又由 (15.37b) 式求得

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \tag{15.38b}$$

代入 $\hat{\tau}_i$ 后, (15.37c) 式可写成

$$\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) - \hat{\beta}\sum_{i=1}^{a}(\bar{x}_{i.} - \bar{x}_{..})\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) + \hat{\beta}S_{xx} = S_{xy}$$

但又注意到

$$\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = T_{xy}$$

$$\sum_{i=1}^{a}(\bar{x}_{i.} - \bar{x}_{..})\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = T_{xx}$$

所以, (15.37c) 式的解为

$$\hat{\beta}=\frac{S_{xy}-T_{xy}}{S_{xx}-T_{xx}}=\frac{E_{xy}}{E_{xx}}$$

这就是前面 15.3.1 节中给出的 (15.26) 式.

拟合全模型 [(15.15) 式] 在总平方和中的减少量可表示为

$$\begin{aligned}R(\mu,\tau,\beta)&=\hat{\mu}y_{..}+\sum_{i=1}^{a}\hat{\tau}_i y_{i.}+\hat{\beta}S_{xy}\\&=(\bar{y}_{..})y_{..}+\sum_{i=1}^{a}[\bar{y}_{i.}-\bar{y}_{..}-(E_{xy}/E_{xx})(\bar{x}_{i.}-\bar{x}_{..})]y_{i.}+(E_{xy}/E_{xx})S_{xy}\\&=y_{..}^2/an+\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})y_{i.}-(E_{xy}/E_{xx})\sum_{i=1}^{a}(\bar{x}_{i.}-\bar{x}_{..})y_{i.}+(E_{xy}/E_{xx})S_{xy}\\&=y_{..}^2/an+T_{yy}-(E_{xy}/E_{xx})(T_{xy}-S_{xy})\\&=y_{..}^2/an+T_{yy}+(E_{xy})^2/E_{xx}\end{aligned}$$

此平方和有 $a+1$ 个自由度, 因为正规方程组的秩是 $a+1$. 此模型的误差平方和是

$$\begin{aligned}SS_E&=\sum_{i=1}^{a}\sum_{j=1}^{n}y_{ij}^2-R(\mu,\tau,\beta)\\&=\sum_{i=1}^{a}\sum_{j=1}^{n}y_{ij}^2-y_{..}^2/an-T_{yy}-(E_{xy})^2/E_{xx}\\&=S_{yy}-T_{yy}-(E_{xy})^2/E_{xx}\\&=E_{yy}-(E_{xy})^2/E_{xx}\end{aligned}\tag{15.39}$$

它有 $an-(a+1)=a(n-1)-1$ 个自由度. 此量即是前面的 (15.27) 式.

今考虑接受零假设, 即 $H_0:\tau_1=\tau_2=\cdots=\tau_a=0$. 此简化模型是

$$y_{ij}=\mu+\beta(x_{ij}-\bar{x}_{..})+\varepsilon_{ij}\quad\begin{cases}i=1,2,\cdots,a\\j=1,2,\cdots,n\end{cases}\tag{15.40}$$

这是一个简单的线性模型, 此模型的最小二乘正规方程组是

$$an\hat{\mu}=y_{..}\tag{15.41a}$$

$$\hat{\beta}S_{xx}=S_{xy}\tag{15.41b}$$

这些方程组的解是 $\hat{\mu}=\bar{y}_{..}$, $\hat{\beta}=S_{xy}/S_{xx}$, 拟合此简化模型而在总平方和中的减少量是

$$R(\mu,\beta)=\hat{\mu}y_{..}+\hat{\beta}S_{xy}=(\bar{y}_{..})y_{..}+(S_{xy}/S_{xx})S_{xy}=y_{..}^2/an+(S_{xy})^2/S_{xx}\tag{15.42}$$

该平方和有两个自由度.

可以发现用于检验 $H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$ 的恰当的平方和是

$$\begin{aligned} R(\tau|\mu,\beta) &= R(\mu,\tau,\beta) - R(\mu,\beta) \\ &= y_{..}^2/an + T_{yy} + (E_{xy})^2/E_{xx} - y_{..}^2/an - (S_{xy})^2/S_{xx} \\ &= S_{yy} - (S_{xy})^2/S_{xx} - [E_{yy} - (E_{xy})^2/E_{xx}] \end{aligned} \tag{15.43}$$

这里用到了 $T_{yy} = S_{yy} - E_{yy}$. 注意 $R(\tau|\mu,\beta)$ 有 $a+1-2=a-1$ 个自由度, 并且等于 15.3.1 节中由 $SS'_E - SS_E$ 给出的平方和. 于是, 针对 $H_0: \tau_i = 0$ 的检验统计量是

$$F_0 = \frac{R(\tau/\mu,\beta)/(a-1)}{SS_E/[a(n-1)-1]} = \frac{(SS'_E - SS_E)/(a-1)}{SS_E/[a(n-1)-1]} \tag{15.44}$$

它即是前面的 (15.30) 式. 因此, 用一般回归显著性检验法, 我们证明了 15.3.1 节的协方差分析的结论.

15.3.4 含协变量的析因实验

协方差分析可以用于更复杂的处理结构, 例如析因设计. 假定对于每种处理组合有足够数据, 则几乎任何复杂的处理结构都可以用协方差分析的方法来分析. 以下要说明如何将协方差分析用于工业实验中最广泛的析因设计类 (2^k 析因设计) 中.

假定在所有处理组合中协变量对响应变量的影响是相同的, 可以算出类似于 15.3.1 节给出的协方差分析表. 唯一的差别是处理平方和. 对于有 n 次重复的 2^2 析因设计, 处理平方和 (T_{yy}) 是 $(1/n)\sum\limits_{i=1}^{2}\sum\limits_{j=1}^{2} y_{ij.}^2 - y_{...}^2/(2)(2)n$. 它是因子 A, B 和交互作用 AB 的平方和的和. 调整的处理平方和可以分解为单个效应部分, 即调整的主效应平方和 SS_A 和 SS_B 以及调整的交互作用的平方和 SS_{AB}.

当扩展处理的设计结构时, 重复的次数是关键问题. 考虑 2^3 析因设计. 若每种处理组合都有单独的协变量时, 要估计出所有的处理组合 (协变量与处理的交互作用), 至少要有两次重复. 这等价于对每个处理组合 (即, 设计单元) 拟合一个简单回归模型. 每个单元有两个观测, 有一个自由度用来估计截距 (处理效应), 另一个自由度来估计斜率 (协变量效应). 对于这种饱和模型, 没有自由度来估计误差. 因此, 一般情况下, 要做完整的协方差分析至少要有 3 次重复. 随着设计单元 (处理组合) 与协变量的数量增加, 此问题变得更加复杂.

若重复的次数有限制, 不同的假定可用来完成某种有用的分析. 可做的最简单的 (通常也是最糟的) 假定是协变量没有效应. 若错误地未考虑协变量, 整个分析与后继的结论也许会有严重的错误. 另一种选择是假定没有处理-协变量交互作用. 即便这个假定是不正确的, 协变量关于所有处理的平均影响, 仍能增加对处理效应的估计与检验的精度. 此假定的一个缺点是, 如果处理的几个水平与协变量有交互作用, 有的项可能会消掉其他项, 而协变量在没有交互作用时给出的估计也会不显著. 第 3 种选择是假定一些因子 (例如某些二因子交互作用或更高阶交互作用) 是不显著的. 这允许一些自由度用于估计误差. 不过, 因为除非有足够的自由度分配给它, 否则误差的估计会相对不精确, 所以这种做法应该很小心地进行, 得到的模型要彻底地评估. 若有两次重复, 上述每种假定都有自由度来估计误差并可以进行有用的假设检验. 采用何种假定由实验情况和实验者愿意承受多大的风险决定. 要注意在建立效应的模型时, 若一个处理因子被去除,

则由原来的 2^3 所得的两次重复不是真正的重复. 这些 "隐性重复" 的确释放了自由度用于参数估计, 但它们不应该作为估计纯误差的重复, 因为原先设计的执行可能没有随机化.

为了说明其中的某些观点, 考虑表 15.16 所示的有两次重复和一个协变量的 2^3 析因设计. 若不考虑协变量来分析响应变量 y, 则相应模型为:

$$\hat{y} = 25.03 + 11.20A + 18.05B + 7.24C - 18.91AB + 14.80AC$$

总模型在 $\alpha = 0.01$ 水平上显著, 而且 $R^2 = 0.786$, $MS_E = 470.82$. 残差分析表明, 除了观测值 $y = 103.01$ 不正常外, 该模型没有问题.

表 15.16　有两次重复的 2^3 设计的响应与协变量数据

A	B	C	x	y
−1	−1	−1	4.05	−30.73
1	−1	−1	0.36	9.07
−1	1	−1	5.03	39.72
1	1	−1	1.96	16.30
−1	−1	1	5.38	−26.39
1	−1	1	8.63	54.58
−1	1	1	4.10	44.54
1	1	1	11.44	66.20
−1	−1	−1	3.58	−26.46
1	−1	−1	1.06	10.94
−1	1	−1	15.53	103.01
1	1	−1	2.92	20.44
−1	−1	1	2.48	−8.94
1	−1	1	13.64	73.72
−1	1	1	−0.67	15.89
1	1	1	5.13	38.57

若选择第 2 种假定, 公共斜率中不包括协变量与处理的交互作用, 则可以估计全部效应与协变量效应. Minitab 输出 (用 General Linear Models 程序) 列在表 15.17 中. 注意, 考虑协变量后, MS_E 大大地减少. 在依次去除各不显著交互效应与主效应 C 后, 最后的分析结果列在表 15.18 中. 此简化模型给出了甚至比表 15.17 中的有协变量的全模型还要小的 MS_E.

最后, 我们考虑第 3 种做法, 假定可以忽略某些交互作用项. 考虑允许处理间以及处理–协变量交互作用的斜率不同的全模型. 假定三因子交互作用 (ABC 与 $ABCx$) 都不显著, 用它们的相应的自由度来估计所拟合的最一般的效应模型中的误差. 这是一个在实践中常用的假定. 在大多数实验中可以忽略三因子以及更高阶的交互作用. Minitab 目前的版本还不能对与处理有交互作用的协变量进行建模, 所以我们采用 SAS PROC GLM. 第III型的平方和是所要的调整的平方和. 表 15.19 给出了此模型的 SAS 结果.

表 15.17 对表 15.16 中的实验进行的 Minitab 协方差分析, 假定有公共斜率

General Linear Model

Factor	Type	Levels	Values
A	fixed	2	-1 1
B	fixed	2	-1 1
C	fixed	2	-1 1

Analysis of Variance for y, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
x	1	12155.9	2521.6	2521.6	28.10	0.001
A	1	1320.7	1403.8	1403.8	15.64	0.005
B	1	3997.6	4066.2	4066.2	45.31	0.000
C	1	52.7	82.3	82.3	0.92	0.370
A*B	1	3788.3	3641.0	3641.0	40.58	0.000
A*C	1	10.2	1.1	1.1	0.01	0.913
B*C	1	5.2	8.4	8.4	0.09	0.769
A*B*C	1	33.2	33.2	33.2	0.37	0.562
Error	7	628.1	628.1	89.7		
Total	15	21992.0				

Term	Coef	StDev	T	P
Constant	-1.016	5.454	-0.19	0.858
x	4.9245	0.9290	5.30	0.001

表 15.18 对表 15.16 中的实验的简化模型进行的 Minitab 协方差分析

General Linear Model

Factor	Type	Levels	Values
A	fixed	2	-1 1
B	fixed	2	-1 1

Analysis of Variance for y, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
x	1	12155.9	8287.9	8287.9	119.43	0.000
A	1	1320.7	1404.7	1404.7	20.24	0.001
B	1	3997.6	4097.7	4097.7	59.05	0.000
A*B	1	3754.5	3754.5	3754.5	54.10	0.000
Error	11	763.3	763.3	69.4		
Total	15	21992.0				

Term	Coef	StDev	T	P
Constant	-1.016	3.225	-0.58	0.572
x	5.0876	0.4655	10.93	0.000

对于一个近乎饱和的模型, 误差估计将会相当不精确. 即便只有很少几项单独在 $\alpha = 0.05$ 水平上显著, 总体感觉此模型比前面两个都要好些 (根据 R^2 与误差均方). 因为对模型的处理效应更感兴趣, 我们依次去除模型的协变量部分中的项, 以增加用于估计误差的自由度. 若依次去除 ACx, BCx, 则 MS_E 会减小到 0.733 6 且有几项不显著. 依次去除 Cx, AC, BC 后, 最终的模型显示在表 15.20 上.

表 15.19　表 15.16 中的实验的 SAS PROC GLM 输出 (协方差分析)

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	21989.20828	1691.47756	1206.45	0.0008
Error	2	2.80406	1.40203		
Corrected Total	15	21992.01234			

R-Square	C.V.	Root MSE	Y Mean
0.999872	4.730820	1.184074	25.02895

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	4.6599694	4.6599694	3.32	0.2099
B	1	13.0525319	13.0525319	9.31	0.0927
C	1	35.0087994	35.0087994	24.97	0.0378
AB	1	17.1013635	17.1013635	12.20	0.0731
AC	1	0.0277472	0.0277472	0.02	0.9010
BC	1	0.4437474	0.4437474	0.32	0.6304
X	1	49.2741287	49.2741287	35.14	0.0273
AX	1	33.9024288	33.9024288	24.18	0.0390
BX	1	92.7747490	95.7747490	68.31	0.0143
CX	1	0.1283784	0.1283784	0.09	0.7908
ABX	1	336.9732676	336.9732676	240.35	0.0041
ACX	1	0.0020997	0.0020997	0.00	0.9726
BCX	1	0.0672386	0.0672386	0.05	0.8470

此例强调, 为了提高关于模型中各项的假设检验的精度, 需要有自由度用于估计实验误差. 过程应该序贯进行, 以避免受不好的误差估计的影响而去除显著的项.

回顾从这 3 种方法得到的结果, 可以看到每种方法都成功地改进了此例中的模型拟合. 若有很强的理由相信协变量与因子没有交互作用, 最好在分析之初就给出假定. 此选择也可以由软件规定. 尽管实验设计软件包也许只能对与处理没有交互作用的协变量建模, 分析者也许仍能合理

选出影响过程的主要因子, 即使存在某个处理–协变量交互作用. 我们也注意到模型适合性的所有常用检验方法都仍然适用, 强烈推荐其作为 ANCOVA 建模过程的组成部分.

表 15.20 表 15.16 中的实验的 SAS PROC GLM 输出 (简化模型)

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	21986.33674	2748.29209	3389.61	0.0001
Error	7	5.67560	0.81080		
Corrected Total	15	21992.01234			

R-Square	C.V.	Root MSE	Y Mean
0.999742	3.597611	0.900444	25.02895

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	19.1597158	19.1597158	23.63	0.0018
B	1	38.0317496	38.0317496	46.91	0.0002
C	1	232.2435668	232.2435668	286.44	0.0001
AB	1	31.7635098	31.7635098	39.18	0.0004
X	1	240.8726525	240.8726525	297.08	0.0001
AX	1	233.3934567	233.3934567	287.86	0.0001
BX	1	550.1530561	550.1530561	678.53	0.0001
ABX	1	542.3268940	542.3268940	668.88	0.0001

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
Intercept	10.2438830	18.74	0.0001	0.54659908
A	2.7850330	4.86	0.0018	0.57291820
B	3.6596279	6.85	0.0002	0.53434356
C	5.4560862	16.92	0.0001	0.32237858
AB	-3.3636850	-6.26	0.0004	0.53741264
X	2.0471937	17.24	0.0001	0.11877417
AX	2.0632049	16.97	0.0001	0.12160595
BX	3.0340997	26.05	0.0001	0.11647826
ABX	-3.0342229	-25.86	0.0001	0.11732045

15.4 重复测量

在社会科学、行为科学以及工程、物理科学的某些方面的实验工作中, 实验的单位经常是人.

在某些实验情况中, 因为经验、所受教育或背景的不同, 不同的人对同一处理的响应的差异是很大的. 除非受到控制, 不然, 人与人之间的这种变异性就会成为实验误差的一部分, 在某些情况下, 它会显著地提高误差均方值, 使得更难于检测出处理之间的真正差异.

利用设计来控制人与人之间的这种变异性是有可能的, 这种设计是对每一个人 (或 "对象") 用 n 种处理的每一种. 这类设计叫做**重复测量设计**. 本节将简要介绍一下单因子重复测量实验.

设一个实验有 a 种处理, 每种处理对 n 位对象的每一位都用一次. 数据如表 15.21 所示. 观测值 y_{ij} 表示对象 j 对处理 i 的响应, 仅用 n 位对象. 用于此设计的模型是

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \tag{15.45}$$

其中 τ_i 是第 i 种处理的效应, β_j 是与第 j 个 "对象" 有关的参数. 假定处理是固定的 (所以 $\sum\limits_{i=1}^{a} \tau_i = 0$), 而所用的对象则是来自作为对象的某一个较大的总体的一个随机样本. 于是, 这些对象总体上代表随机效应, 所以我们假定 β_j 的均值是零, β_j 的方差是 σ_β^2. 因为项 β_j 对同一位对象的所有 a 次测量是公用的, 所以 y_{ij} 与 $y_{i'j}$ 之间的协方差一般说来不是零. 习惯上假定 y_{ij} 与 $y_{i'j}$ 之间的协方差对所有的处理和对象是常数.

表 15.21 一个单因子重复测量设计的数据

处理	对象				处理总和
	1	2	$\cdots$	n	
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
a	y_{a1}	y_{a2}		y_{an}	$y_{a.}$
对象总和	$y_{.1}$	$y_{.2}$	$\cdots$	$y_{.n}$	$y_{..}$

考虑总平方和的方差分解式的一种分析法, 比方说

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = a\sum_{j=1}^{n}(\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{.j})^2 \tag{15.46}$$

可以将 (15.46) 式右边的第 1 项看作是由对象之间差异所得的平方和, 第 2 项看作是对象内部差异的平方和, 也就是说,

$$SS_T = SS_{\text{对象之间}} + SS_{\text{对象内部}}$$

平方和 $SS_{\text{对象之间}}$ 与 $SS_{\text{对象内部}}$ 是统计独立的, 其自由度为

$$an - 1 = (n-1) + n(a-1)$$

对象内部的差异既依赖于处理效应的差异, 也依赖于不可控制的变异性 (噪音或误差). 因此, 可以将由对象内部差异所得的平方和分解为

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{.j})^2 = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \tag{15.47}$$

(15.47) 式右边第 1 项测量了处理均值之间的差异对 $SS_{对象内部}$ 的贡献, 而第 2 项是属于误差的残差偏差. $SS_{对象内部}$ 的两个分量是独立的. 于是

$$SS_{对象内部} = SS_{处理} + SS_E$$

其自由度分别为

$$n(a-1) = (a-1) + (a-1)(n-1).$$

为检验没有处理效应的假设, 即,

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$$

$$H_1: 至少有一个\ \tau_i \neq 0$$

用比值

$$F_0 = \frac{SS_{处理}/(a-1)}{SS_E/(a-1)(n-1)} = \frac{MS_{处理}}{MS_E} \tag{15.48}$$

如果模型误差是正态分布的, 则在零假设 $H_0: \tau_i = 0$ 下, 统计量 F_0 服从 $F_{a-1,(a-1)(n-1)}$ 分布. 当 $F_0 > F_{\alpha,a-1,(a-1)(n-1)}$ 时拒绝零假设.

方差分析法概括在表 15.22 中, 表中还列出了便于计算的平方和公式. 读者应该能认出, 重复测量的单因子设计的方差分析等价于随机化完全区组设计的分析, 将对象看作为区组就是了.

表 15.22 单因子重复度量设计的方差分析

方差来源	平方和	自由度	均 方	F_0
(1) 对象之间	$\sum_{j=1}^{n} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{an}$	$n-1$		
(2) 对象内部	$\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \sum_{j=1}^{n} \frac{y_{.j}^2}{a}$	$n(a-1)$		
(3) (处理)	$\sum_{i=1}^{a} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an}$	$a-1$	$MS_{处理} = \frac{SS_{处理}}{a-1}$	$\frac{MS_{处理}}{MS_E}$
(4) (误差)	相减: 行 (2)– 行 (3)	$(a-1)(n-1)$	$MS_E = \frac{SS_E}{(a-1)(n-1)}$	
(5) 总和	$\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{an}$	$an-1$		

15.5 思 考 题

*15.1 重新考虑思考题 5.22 中的实验. 用 Box-Cox 法确定, 在该实验数据的分析中, 对响应进行变换是否恰当 (或有用)?

*15.2 在例 6.3 中, 我们选择了对钻头推进速率响应做对数变换. 用 Box-Cox 法说明这是恰当的数据变换.

15.3 重新考虑思考题 8.24 中的熔炼工序实验, 该实验用一个 2^{6-3} 分式析因设计研究烘烤后粘在阳极上的粘贴材料的重量. 对设计的 8 次试验的每一次重复 3 次, 每个试验组合的重量的平均值与极差作为响应变量. 是否有迹象表明需要对各响应变量做变换?

15.4 思考题 8.25 中, 一个重复分式析因设计用于研究半导体制造中的基层的弯曲度. 基层的弯曲度测量值的均值与标准差作为响应变量. 是否有迹象表明需要对各响应变量做变换?

15.5 重新考虑思考题 8.26 中的光刻胶实验. 把每个试验组合的光刻胶厚度的方差作为响应变量. 是否有迹象表明需要对各响应变量做变换?

*15.6 在思考题 8.30 描述的护栅缺陷的实验中，一种平方根变换的变形被用于数据分析. 用 Box-Cox 法确定这是否恰当的数据变换.

15.7 在思考题 12.11 的中心复合设计中，得到两个响应：氧化物厚度的均值与方差. 用 Box-Cox 法研究变换对这两个响应变量的用处. 该问题的 (c) 部分中提出的对数变换是否恰当?

15.8 在思考题 12.12 的 3^3 分式析因设计中，有一个响应变量是标准差. 用 Box-Cox 法研究变换对这两个响应变量的用处. 若以方差作为响应变量，结果是否不同?

15.9 思考题 12.9 建议用 $\ln(s^2)$ 作为响应变量 [根据 (b) 部分]. Box-Cox 法显示此变换是否合适?

15.10 Myers, Montgomery, and Vining(2002) 描述了一个研究精子生存的实验. 设计因子是柠檬酸钠的用量、丙三醇的用量以及平衡时间，每个因子都有两种水平. 响应变量是在各组条件安排下每 50 个精子中存活的精子数量. 数据见下表：

柠檬酸钠	丙三醇	平衡时间	存活数量
−	−	−	34
+	−	−	20
−	+	−	8
+	+	−	21
−	−	+	30
+	−	+	20
−	+	+	10
+	+	+	25

用 logistic 回归法分析该实验数据.

15.11 一个软饮料批发商研究运输方法的效果. 开发了 3 种手推车，在公司的方法工程实验室进行了一个实验. 感兴趣的变量是运输时间 (y, 以分钟为单位). 不过，运输时间也与运送物体积 (x) 有很强的关系. 每种手推车用 4 次，得到下面的数据. 分析这些数据并得出合适的结论. (取 $\alpha = 0.05$)

手推车类型					
1		2		3	
y	x	y	x	y	x
27	24	25	26	40	38
44	40	35	32	22	26
33	35	46	42	53	50
41	40	26	25	18	20

15.12 对于思考题 15.11 中的数据，计算调整的处理均值及其标准误.

15.13 下面是单因子协方差分析的平方和与叉积和. 完成分析并作合适的结论. (取 $\alpha = 0.05$)

方差来源	自由度	平方和与叉积和		
		x	xy	y
处理	3	1 500	1 000	650
误差	12	6 000	1 200	550
总和	15	7 500	2 200	1 200

15.14 求例 15.5 中调整的处理均值的标准误.

*15.15 测试某种工业胶水的 4 种配方. 胶水用在粘合部件时的拉伸强度与所用厚度有关. 对各配方, 得到强度 (y) 的 5 个观测值 (以磅为单位) 和厚度 (x, 以 0.01 英寸为单位). 数据列在下表. 分析这些数据并得出合适的结论.

胶水配方							
1		2		3		4	
y	x	y	x	y	x	y	x
46.5	13	48.7	12	46.3	15	44.7	16
45.9	14	49.0	10	47.1	14	43.0	15
49.8	12	50.1	11	48.9	11	51.0	10
46.1	12	48.5	12	48.2	11	48.1	12
44.3	14	45.2	14	50.3	10	48.6	11

15.16 使用思考题 15.11 的数据, 计算调整的处理均值与它们的标准误.

15.17 一名工程师正研究机器操作时切割速度对金属切削率的效应. 不过, 金属切削率也与试样的硬度有关. 每种切割速度取 5 个观测值. 被切削的金属量 (y) 与试样的硬度 (x) 列在下表中. 用协方差分析法分析此数据. (取 $\alpha = 0.05$)

切割速度 (转/分钟)					
1 000		1 200		1 400	
y	x	y	x	y	x
68	120	112	165	118	175
90	140	94	140	82	132
98	150	65	120	73	124
77	125	74	125	92	141
88	136	85	133	80	130

15.18 证明: 在有单个协变量的单因子协方差分析中, 第 i 个调整的处理均值的 $100(1-\alpha)\%$ 置信区间为

$$\bar{y}_{i.} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \pm \tau_{\alpha/2,a(n-1)-1}\left[MS_E\left(\frac{1}{n} + \frac{(\bar{x}_{i.} - \bar{x}_{i..})^2}{E_{xx}}\right)\right]^{1/2}$$

用此公式, 计算例 15.5 中机器 1 的调整的处理均值的 95%置信区间.

15.19 证明: 在有单个协变量的单因子协方差分析中, 任意两个调整的处理均值的差是

$$S_{\text{调整的}\bar{y}_{i.}-\text{调整的}\bar{y}_{j.}} = \left[MS_E\left(\frac{2}{n} + \frac{(\bar{x}_{i.} - \bar{x}_{j.})^2}{E_{xx}}\right)\right]^{1/2}$$

15.20 试讨论如何将方差分析中的抽检特性曲线用于协方差分析.

附　　录

I. 累积标准正态分布
II. t 分布的百分位数
III. χ^2 分布的百分位数
IV. F 分布的百分位数
V. 固定效应模型方差分析的抽检特性曲线
VI. 随机效应模型方差分析的抽检特性曲线
VII. 学生化极差统计量的百分位数
VIII. 带有控制的比较处理的 Dunnett 检验的临界值
IX. 正交多项式的系数
X. 2^{k-p} 分式析因设计的别名关系 ($k \leqslant 15, n \leqslant 64$)

I. 累积标准正态分布①

$$\Phi(z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-u^2/2}\mathrm{d}u$$

z	0.00	0.01	0.02	0.03	0.04	z
0.0	0.500 00	0.503 99	0.507 98	0.511 97	0.515 95	0.0
0.1	0.539 83	0.543 79	0.547 76	0.551 72	0.555 67	0.1
0.2	0.579 26	0.583 17	0.587 06	0.590 95	0.594 83	0.2
0.3	0.617 91	0.621 72	0.625 51	0.629 30	0.633 07	0.3
0.4	0.655 42	0.659 10	0.662 76	0.666 40	0.670 03	0.4
0.5	0.691 46	0.694 97	0.698 47	0.701 94	0.705 40	0.5
0.6	0.725 75	0.729 07	0.732 37	0.735 65	0.738 91	0.6
0.7	0.758 03	0.761 15	0.764 24	0.767 30	0.770 35	0.7
0.8	0.788 14	0.791 03	0.793 89	0.796 73	0.799 54	0.8
0.9	0.815 94	0.818 59	0.821 21	0.823 81	0.826 39	0.9
1.0	0.841 34	0.843 75	0.846 31	0.848 49	0.850 83	1.0
1.1	0.864 33	0.866 50	0.868 64	0.870 76	0.872 85	1.1
1.2	0.884 93	0.886 86	0.888 77	0.890 65	0.892 51	1.2
1.3	0.903 20	0.904 90	0.906 58	0.908 24	0.909 88	1.3
1.4	0.919 24	0.920 73	0.922 19	0.923 64	0.925 06	1.4
1.5	0.933 19	0.934 48	0.935 74	0.936 99	0.938 22	1.5
1.6	0.945 20	0.946 30	0.947 38	0.948 45	0.949 50	1.6
1.7	0.955 43	0.956 37	0.957 28	0.958 18	0.959 07	1.7
1.8	0.964 07	0.964 85	0.965 62	0.966 37	0.967 11	1.8
1.9	0.971 28	0.971 93	0.972 57	0.973 20	0.973 81	1.9
2.0	0.977 25	0.977 78	0.978 31	0.978 82	0.979 32	2.0
2.1	0.982 14	0.982 57	0.983 00	0.983 41	0.938 82	2.1
2.2	0.986 10	0.986 45	0.986 79	0.987 13	0.987 45	2.2
2.3	0.989 28	0.989 56	0.989 83	0.990 10	0.990 36	2.3
2.4	0.991 80	0.992 02	0.992 24	0.992 45	0.992 66	2.4
2.5	0.993 79	0.993 96	0.994 13	0.994 30	0.994 46	2.5
2.6	0.995 34	0.995 47	0.995 60	0.995 73	0.995 85	2.6
2.7	0.996 53	0.996 64	0.996 74	0.996 83	0.996 93	2.7
2.8	0.997 44	0.997 52	0.997 60	0.997 67	0.997 74	2.8
2.9	0.998 13	0.998 19	0.998 25	0.998 31	0.998 36	2.9
3.0	0.998 65	0.998 69	0.998 74	0.998 78	0.998 82	3.0
3.1	0.999 03	0.999 06	0.999 10	0.999 13	0.999 16	3.1
3.2	0.999 31	0.999 34	0.999 36	0.999 38	0.999 40	3.2
3.3	0.999 52	0.999 53	0.999 55	0.999 57	0.999 58	3.3
3.4	0.999 66	0.999 68	0.999 69	0.999 70	0.999 71	3.4
3.5	0.999 77	0.999 78	0.999 78	0.999 79	0.999 80	3.5
3.6	0.999 84	0.999 85	0.999 85	0.999 86	0.999 86	3.6
3.7	0.999 89	0.999 90	0.999 90	0.999 90	0.999 91	3.7
3.8	0.999 93	0.999 93	0.999 93	0.999 94	0.999 94	3.8
3.9	0.999 95	0.999 95	0.999 96	0.999 96	0.999 96	3.9

① 承蒙惠允, 摘自 W. W. Hines and D. C. Montgomery, *Probability and Statistics in Engineering and Management Science*, 3rd edition, Wiley, New York, 1990.

(续)

z	0.05	0.06	0.07	0.08	0.09	z
0.0	0.519 94	0.523 92	0.527 90	0.531 88	0.535 86	0.0
0.1	0.559 62	0.563 56	0.567 49	0.571 42	0.575 34	0.1
0.2	0.598 71	0.602 57	0.606 42	0.610 26	0.614 09	0.2
0.3	0.636 83	0.640 58	0.644 31	0.648 03	0.651 73	0.3
0.4	0.673 64	0.677 24	0.680 82	0.684 38	0.687 93	0.4
0.5	0.708 84	0.712 26	0.715 66	0.719 04	0.722 40	0.5
0.6	0.742 15	0.745 37	0.748 57	0.751 75	0.754 90	0.6
0.7	0.773 37	0.776 37	0.779 35	0.782 30	0.785 23	0.7
0.8	0.802 34	0.805 10	0.807 85	0.810 57	0.813 27	0.8
0.9	0.828 94	0.831 47	0.833 97	0.836 46	0.838 91	0.9
1.0	0.853 14	0.855 43	0.857 69	0.859 93	0.862 14	1.0
1.1	0.874 93	0.876 97	0.879 00	0.881 00	0.882 97	1.1
1.2	0.894 35	0.896 16	0.897 96	0.899 73	0.901 47	1.2
1.3	0.911 49	0.913 08	0.914 65	0.916 21	0.917 73	1.3
1.4	0.926 47	0.927 85	0.929 22	0.930 56	0.931 89	1.4
1.5	0.939 43	0.904 62	0.941 79	0.942 95	0.944 08	1.5
1.6	0.950 53	0.951 54	0.952 54	0.953 52	0.954 48	1.6
1.7	0.959 94	0.960 80	0.961 64	0.962 46	0.963 27	1.7
1.8	0.967 84	0.968 56	0.969 26	0.969 95	0.970 62	1.8
1.9	0.974 41	0.975 00	0.975 58	0.976 15	0.976 70	1.9
2.0	0.979 82	0.980 30	0.980 77	0.981 24	0.981 69	2.0
2.1	0.984 22	0.984 61	0.985 00	0.985 37	0.985 74	2.1
2.2	0.987 78	0.988 09	0.988 40	0.988 70	0.988 99	2.2
2.3	0.990 61	0.990 86	0.991 11	0.991 34	0.991 58	2.3
2.4	0.992 86	0.993 05	0.993 24	0.993 43	0.993 61	2.4
2.5	0.994 61	0.994 77	0.994 92	0.995 06	0.995 20	2.5
2.6	0.995 98	0.996 09	0.996 21	0.996 32	0.996 43	2.6
2.7	0.997 02	0.997 11	0.997 20	0.997 28	0.997 36	2.7
2.8	0.997 81	0.997 88	0.997 95	0.998 01	0.998 07	2.8
2.9	0.998 41	0.998 46	0.998 51	0.998 56	0.998 61	2.9
3.0	0.998 86	0.998 89	0.998 93	0.998 97	0.999 00	3.0
3.1	0.999 18	0.999 21	0.999 24	0.999 26	0.999 29	3.1
3.2	0.999 42	0.999 44	0.999 46	0.999 48	0.999 50	3.2
3.3	0.999 60	0.999 61	0.999 62	0.999 64	0.999 65	3.3
3.4	0.999 72	0.999 73	0.999 74	0.999 75	0.999 76	3.4
3.5	0.999 81	0.999 81	0.999 82	0.999 83	0.999 83	3.5
3.6	0.999 87	0.999 87	0.999 88	0.999 88	0.999 89	3.6
3.7	0.999 91	0.999 92	0.999 92	0.999 92	0.999 92	3.7
3.8	0.999 94	0.999 94	0.999 95	0.999 95	0.999 95	3.8
3.9	0.999 96	0.999 96	0.999 96	0.999 97	0.999 97	3.9

II. t 分布的百分位点 ①

ν \ α	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.002 5	0.001	0.000 5
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.713	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.019	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

ν = 自由度

① 承蒙惠允，摘自 E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Vol. 1, 3rd edition, Cambridge University Press, Cambridge, 1966.

III. χ^2 分布的百分位点[①]

ν \ α	0.995	0.990	0.975	0.950	0.500	0.050	0.025	0.010	0.005
1	0.00+	0.00+	0.00+	0.00+	0.45	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	1.39	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	2.37	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	3.36	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	4.35	11.07	12.38	15.09	16.75
6	0.68	0.87	1.24	1.64	5.35	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	6.35	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	7.34	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	8.34	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	9.34	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	10.34	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	11.34	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	12.34	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	13.34	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	14.34	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	15.34	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	16.34	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	17.34	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	18.34	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	19.34	31.41	34.17	37.57	40.00
25	10.52	11.52	13.12	14.61	24.34	37.65	40.65	44.31	46.93
30	13.79	14.95	16.79	18.49	29.34	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	39.34	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	49.33	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	59.33	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	69.33	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	79.33	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	89.33	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	99.33	124.34	129.56	135.81	140.17

ν = 自由度

① 承蒙惠允，摘自 E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Vol. 1, 3rd edition, Cambridge University Press, Cambridge, 1966.

IV. F 分布的百分位点①

$F_{0.25,\nu_1,\nu_2}$

ν_2 \ ν_1 分子自由度 (ν_1)	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.41	9.49	9.58	9.63	9.67	9.71	9.76	9.80	9.85
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48
3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87	1.87	1.87
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.76	1.76	1.75	1.75	1.75	1.74	1.74	1.74
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.70	1.69	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.62	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.53	1.53
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.53	1.52	1.52	1.51	1.51	1.50	1.49	1.48
11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47	1.46	1.45
12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.42
分 13	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40
母 14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.41	1.40	1.39	1.38
自 15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.43	1.41	1.41	1.40	1.39	1.38	1.37	1.36
由 16	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.44	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
度 17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33
ν_2 18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32
19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.30
20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29
21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28
22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.37	1.36	1.34	1.33	1.32	1.31	1.30	1.29	1.28
23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28	1.27
24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.29	1.28	1.26
25	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.32	1.31	1.29	1.28	1.27	1.25
26	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.35	1.34	1.32	1.31	1.30	1.29	1.28	1.26	1.25
27	1.38	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.33	1.32	1.31	1.30	1.28	1.27	1.26	1.24
28	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.31	1.30	1.29	1.28	1.27	1.25	1.24
29	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.32	1.31	1.30	1.29	1.27	1.26	1.25	1.23
30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.34	1.32	1.30	1.29	1.28	1.27	1.26	1.24	1.23
40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.31	1.30	1.28	1.26	1.25	1.24	1.22	1.21	1.19
60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.27	1.25	1.24	1.22	1.21	1.19	1.17	1.15
120	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.26	1.24	1.22	1.21	1.19	1.18	1.16	1.13	1.10
∞	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.22	1.19	1.18	1.16	1.14	1.12	1.08	1.00

ν = 自由度

① 承蒙惠允，摘自 E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Vol. 1, 3rd edition, Cambridge University Press, Cambridge, 1966.

$F_{0.10,\nu_1,\nu_2}$

(续)

ν_2 \ ν_1 (分母自由度 ν_2 / 分子自由度 ν_1)	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.96	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.03	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

$F_{0.05,\nu_1,\nu_2}$

(续)

ν_2 \ ν_1	分子自由度 (ν_1)																	
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
分 13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
母 14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
自 15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
由 16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
度 17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
(ν_2) 18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.96	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.55	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

$F_{0.025,\nu_1,\nu_2}$ (续)

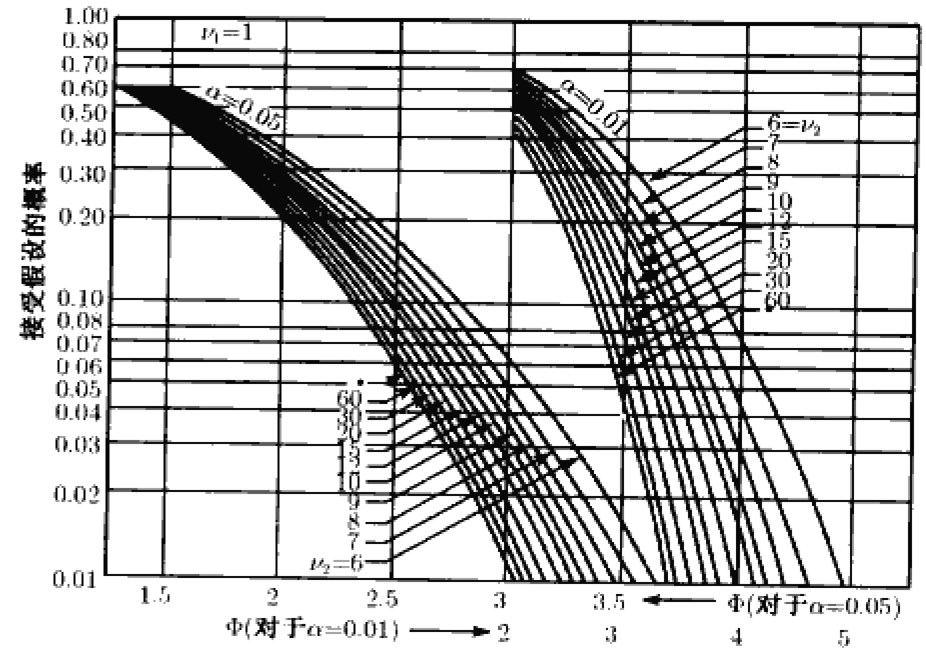
ν_2 \ ν_1 分子自由度 (ν_1)	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1 001	1 006	1 010	1 014	1 018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.76	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
分 13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
母 14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
自 15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
由 16	6.12	2.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
度 17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
(ν_2) 18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	1.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.02	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

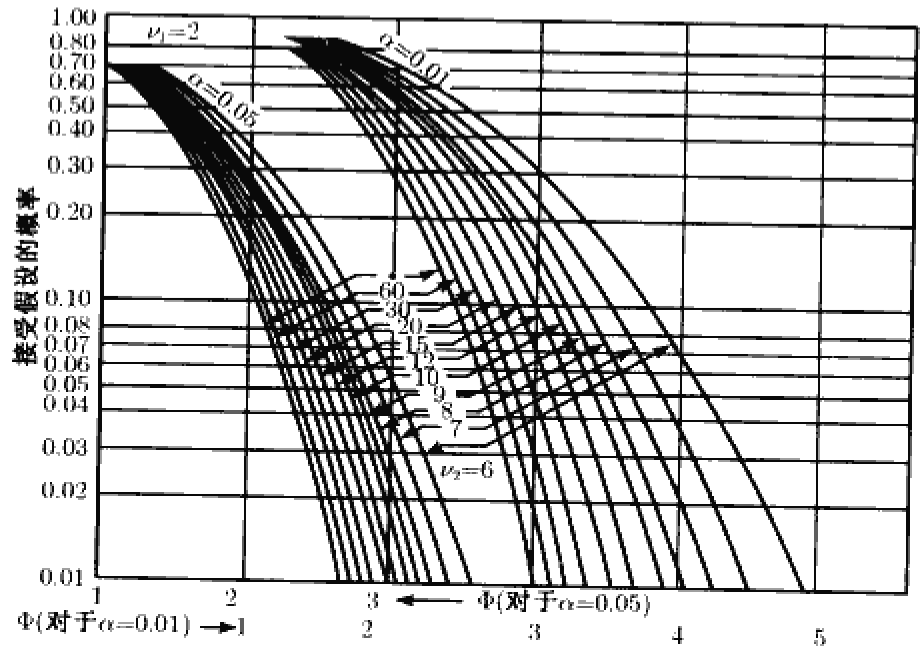
$F_{0.01,\nu_1,\nu_2}$

(续)

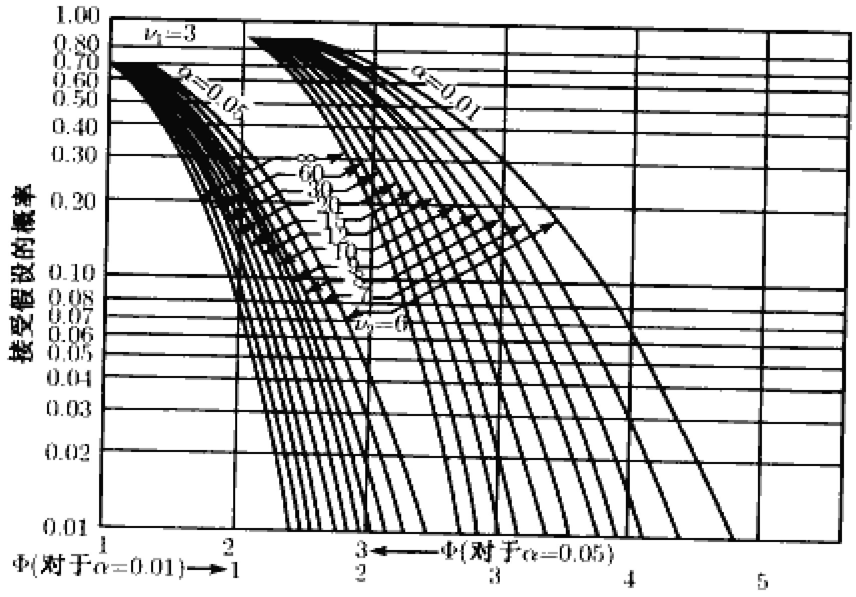
ν_2 \ ν_1	分子自由度 (ν_1)																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4 052	4 999.5	5 403	5 625	5 764	5 859	5 928	5 982	6 022	6 056	6 106	6 157	6 209	6 235	6 261	6 287	6 313	6 339	6 366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.96	26.00	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
分 13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
母 14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
自 15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
由 16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
度 17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
(ν_2) 18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.59	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.34	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

V. 固定效应模型方差分析的抽检特性曲线①

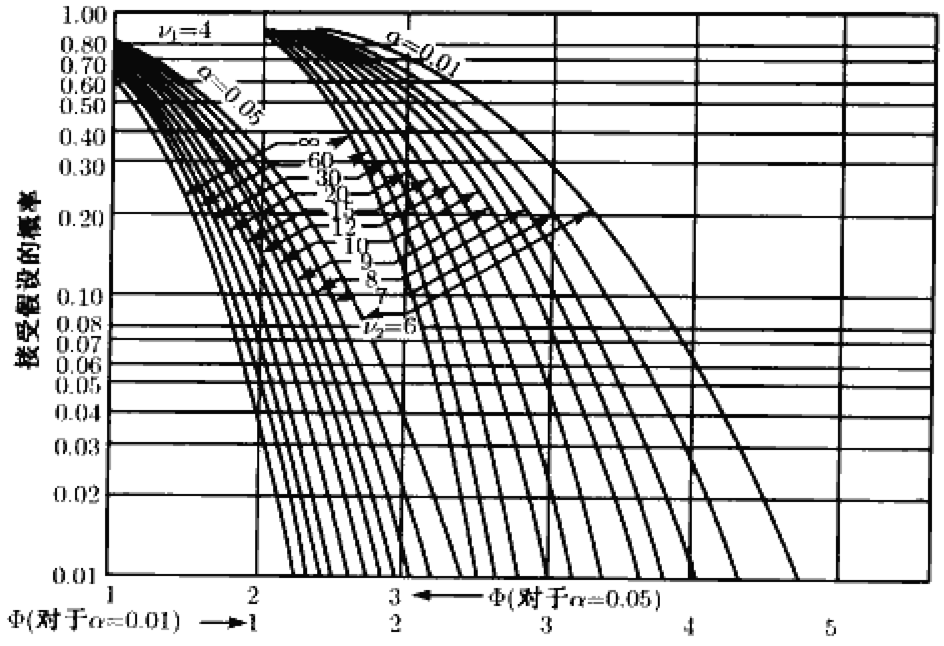


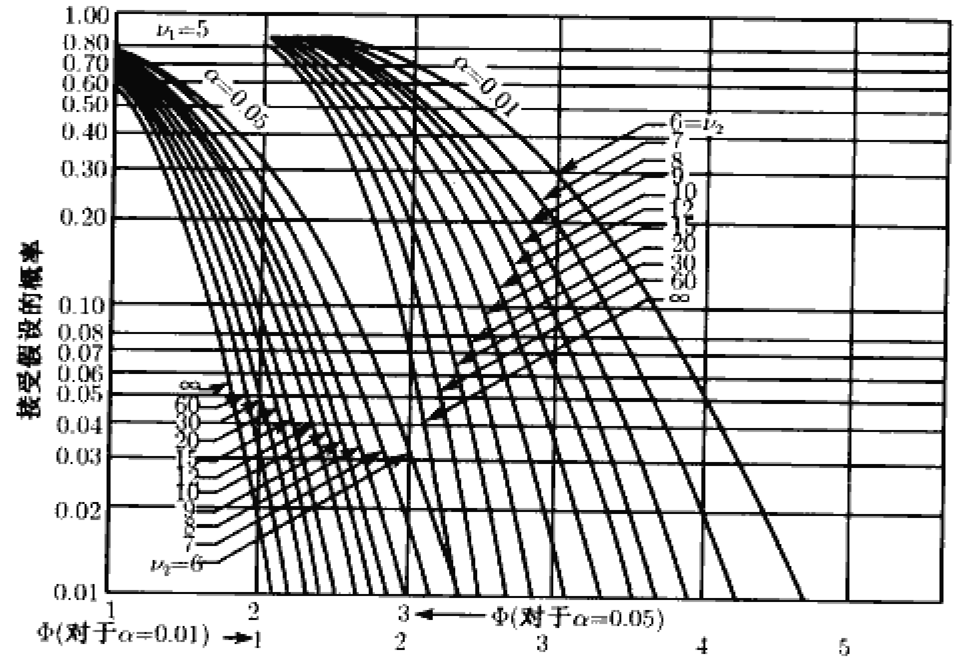


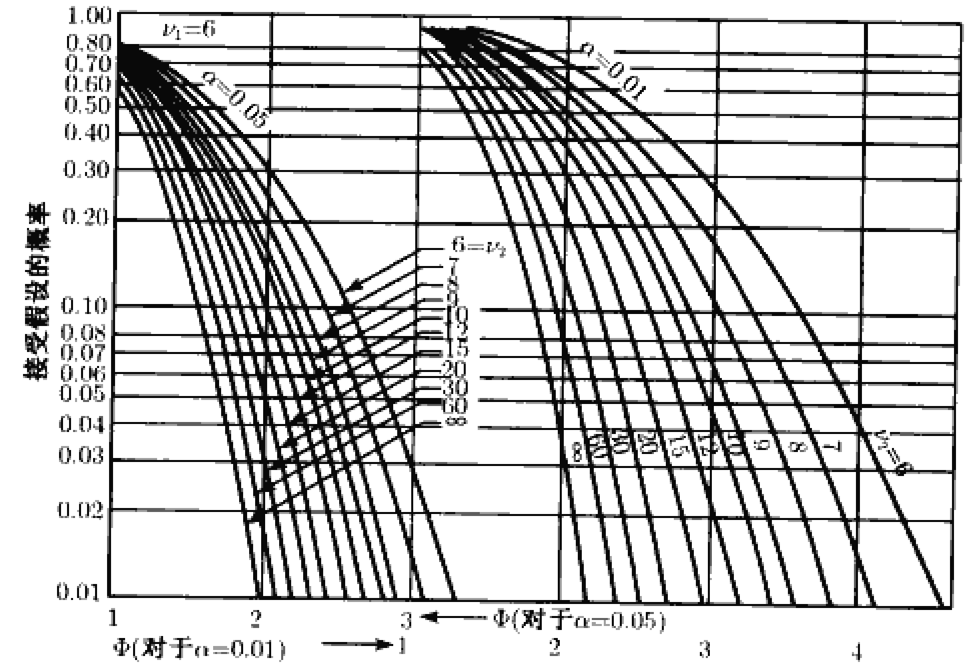
ν_1=分子自由度, ν_2=分母自由度

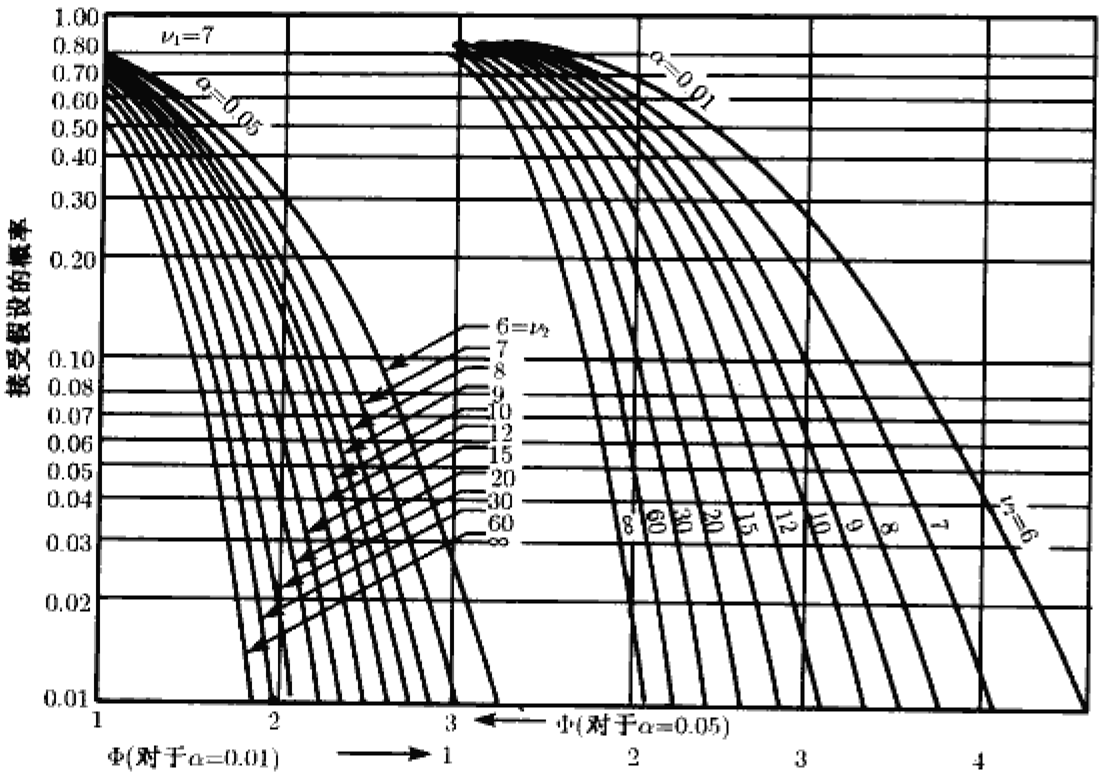


① 承蒙惠允, 摘自 E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Vol. 2, Cambridge University Press, Cambridge, 1972.

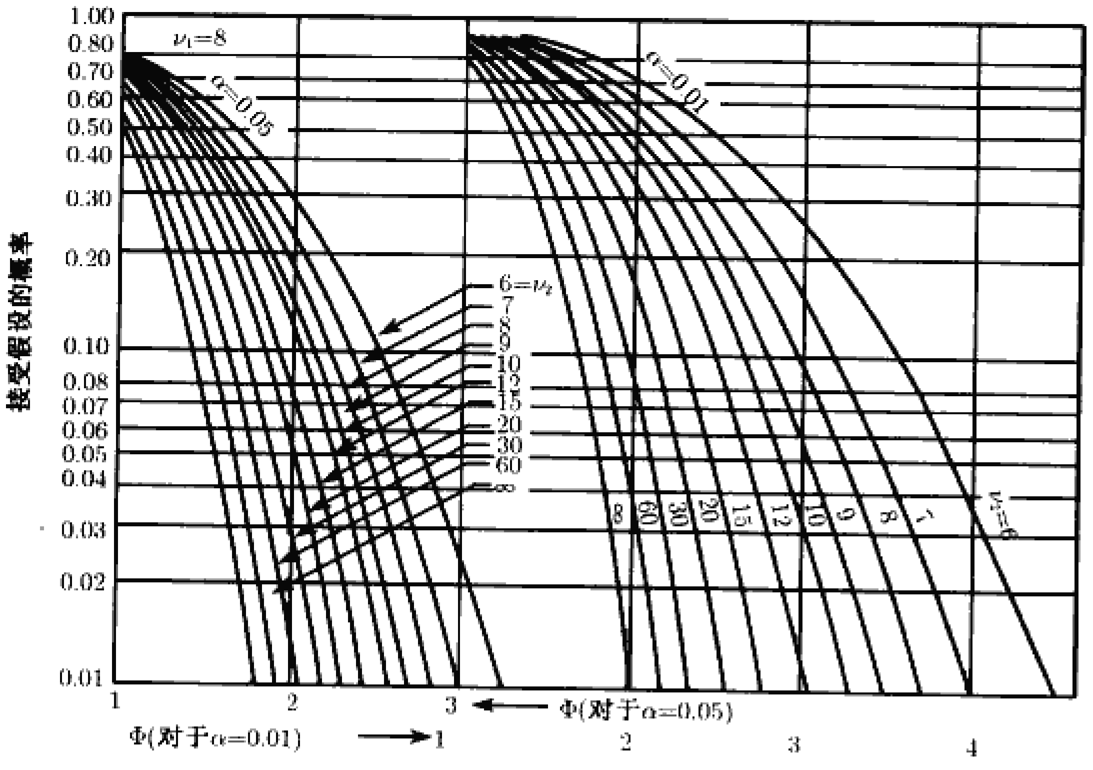

$\nu_1=4$
$\alpha=0.05$
$\alpha=0.01$
$\nu_2=6$
接受假设的概率
Φ(对于α=0.05)
Φ(对于α=0.01)


$\nu_1=5$
$\alpha=0.05$
$\alpha=0.01$
$6=\nu_2$
$\nu_2=6$
接受假设的概率
Φ(对于α=0.05)
Φ(对于α=0.01)


$\nu_1=6$
$\alpha=0.05$
$\alpha=0.01$
$6=\nu_2$
$\nu_2=6$
接受假设的概率
Φ(对于α=0.05)
Φ(对于α=0.01)

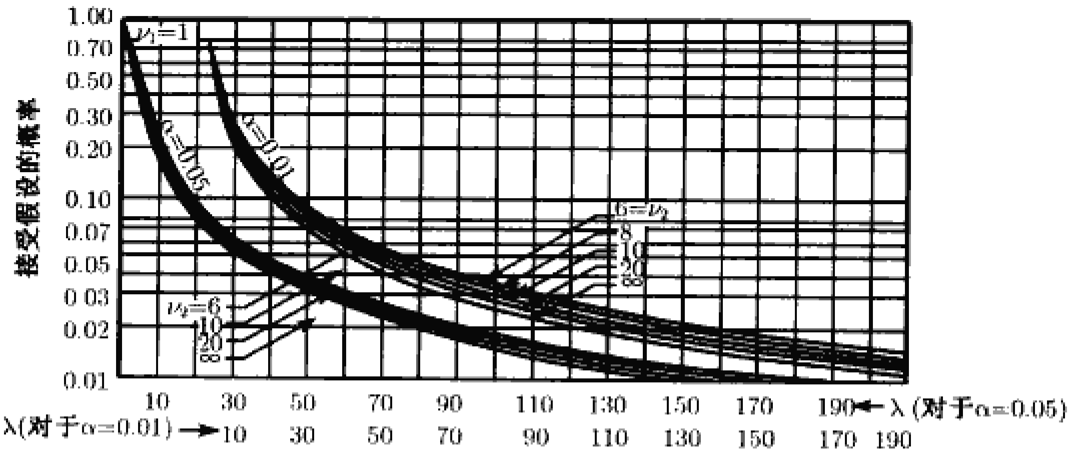


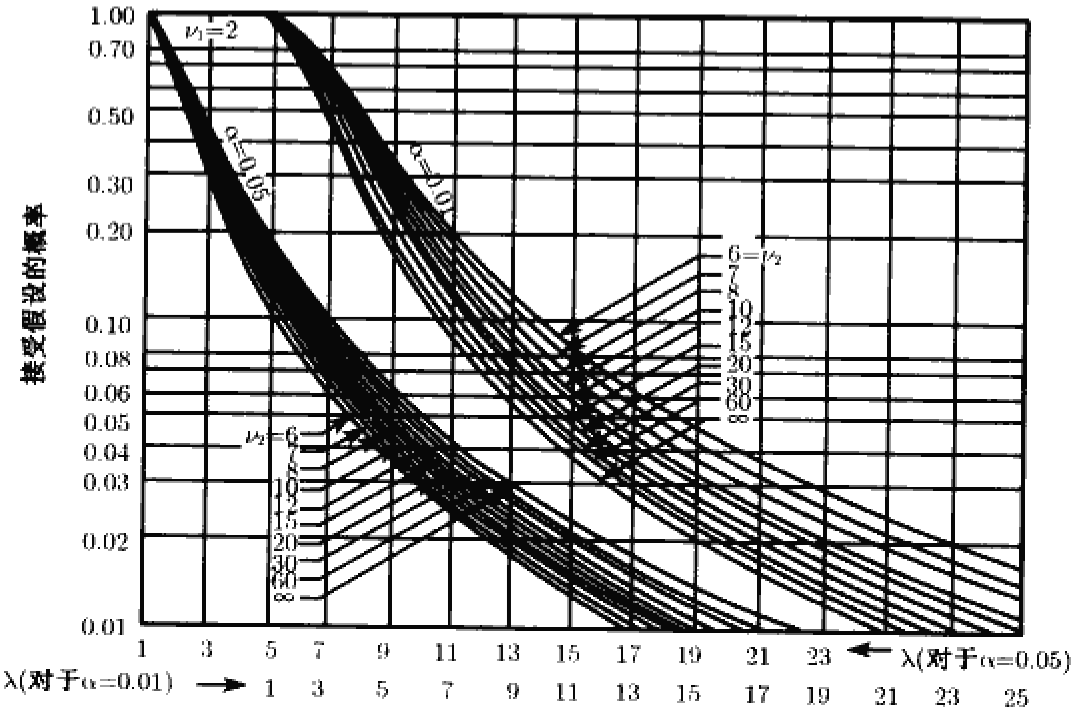
接受假设的概率
$\nu_1=7$
$\alpha=0.05$
$\alpha=0.01$
$6=\nu_2$
$\nu_2=6$
Φ(对于α=0.01)
Φ(对于α=0.05)

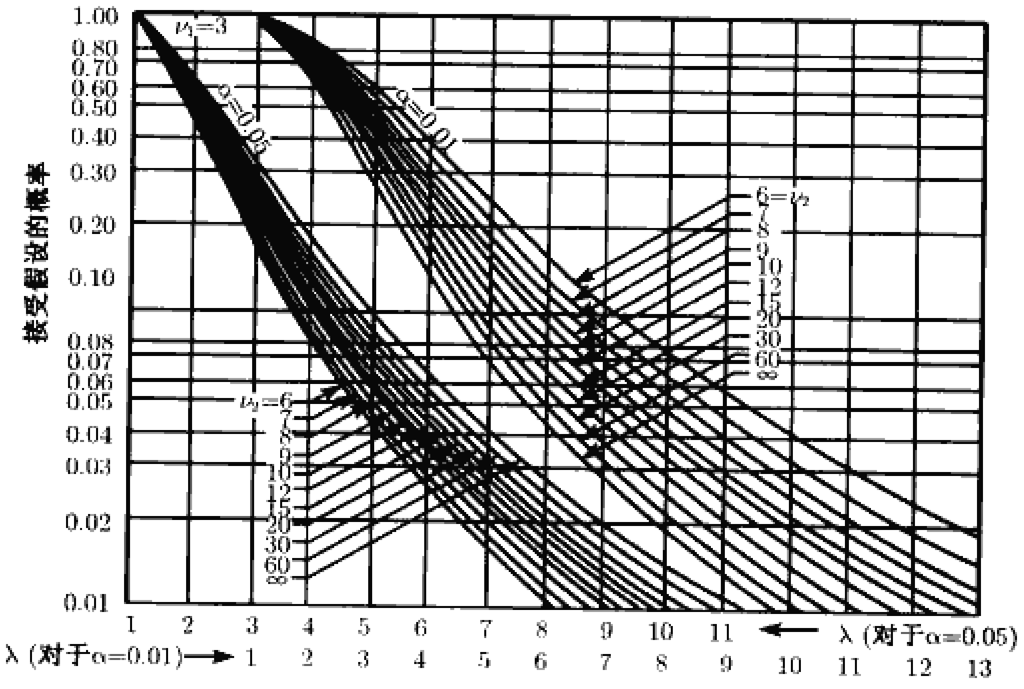


接受假设的概率
$\nu_1=8$
$\alpha=0.05$
$\alpha=0.01$
$6=\nu_2$
$\nu_2=6$
Φ(对于α=0.01)
Φ(对于α=0.05)

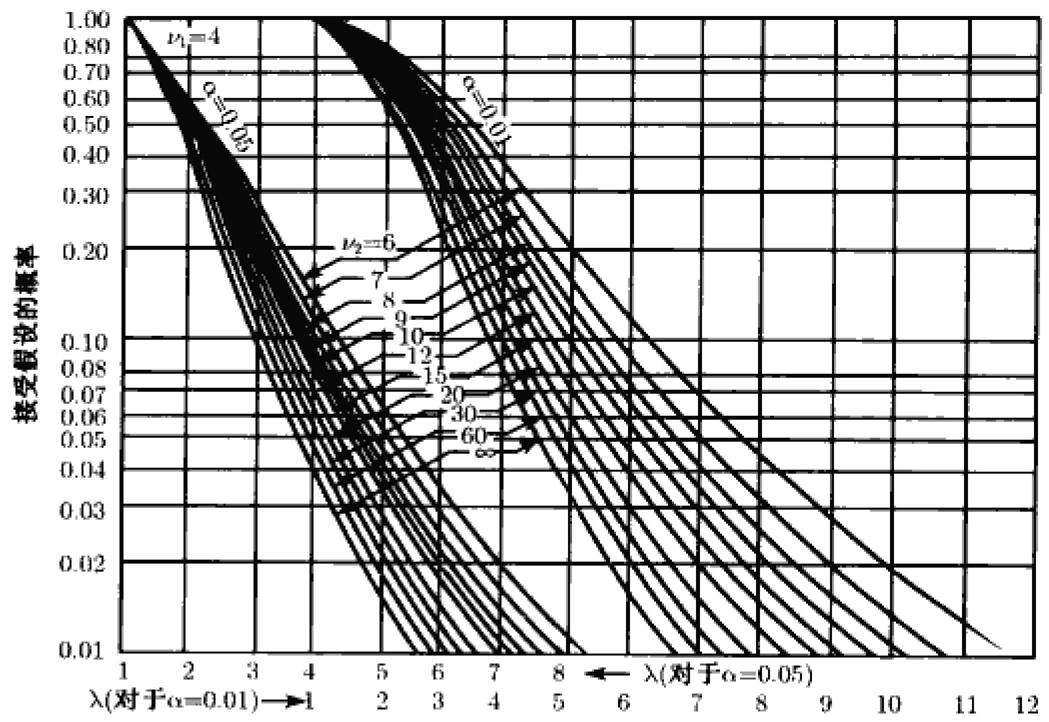
VI. 随机效应模型方差分析的抽检特性曲线 ①



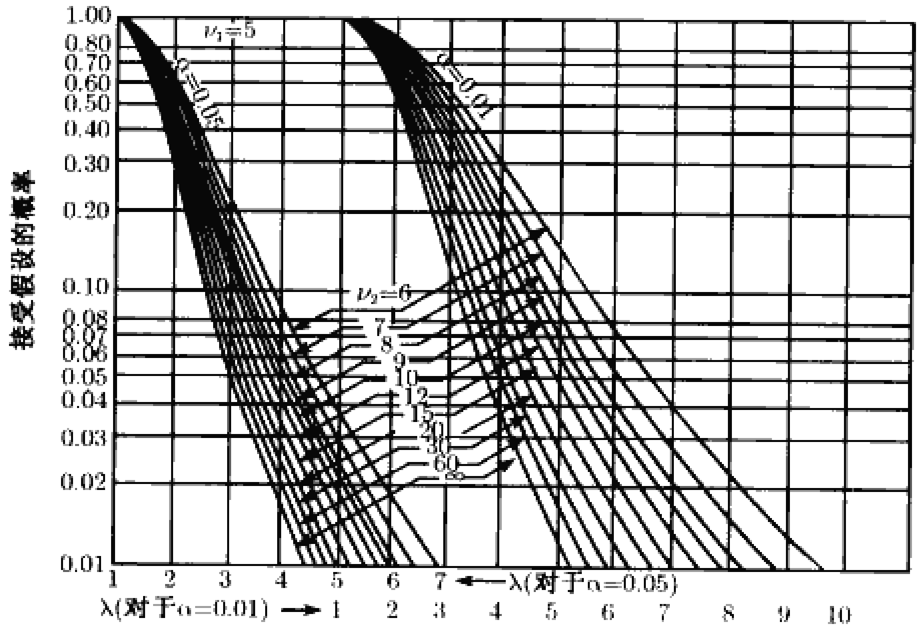




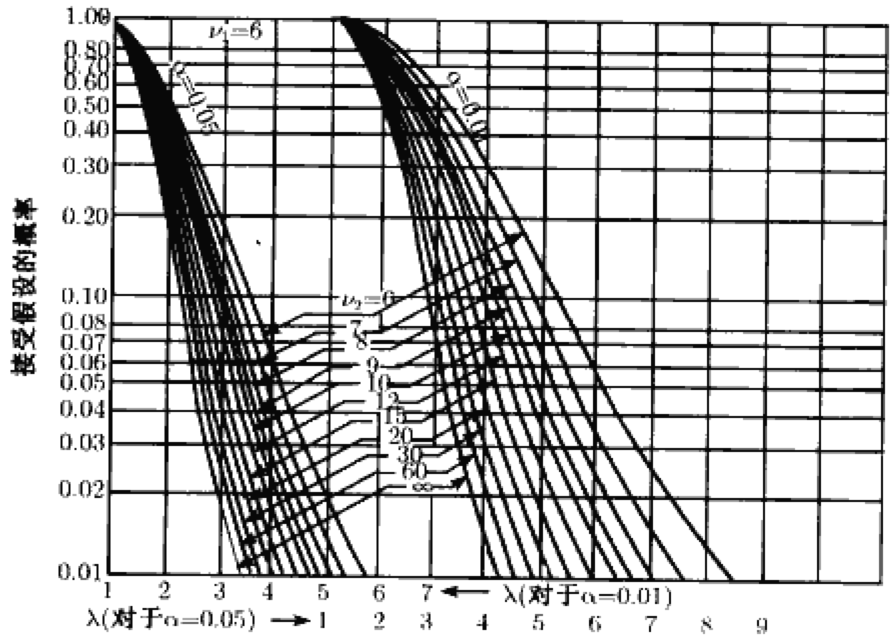
① 承蒙惠允，摘自 A. H. Bowker and G. J. Lieberman, *Engineering Statistics*, 2nd edition, Prentice Hall, Inc., Englewood Cliffs, N. J., 1972.



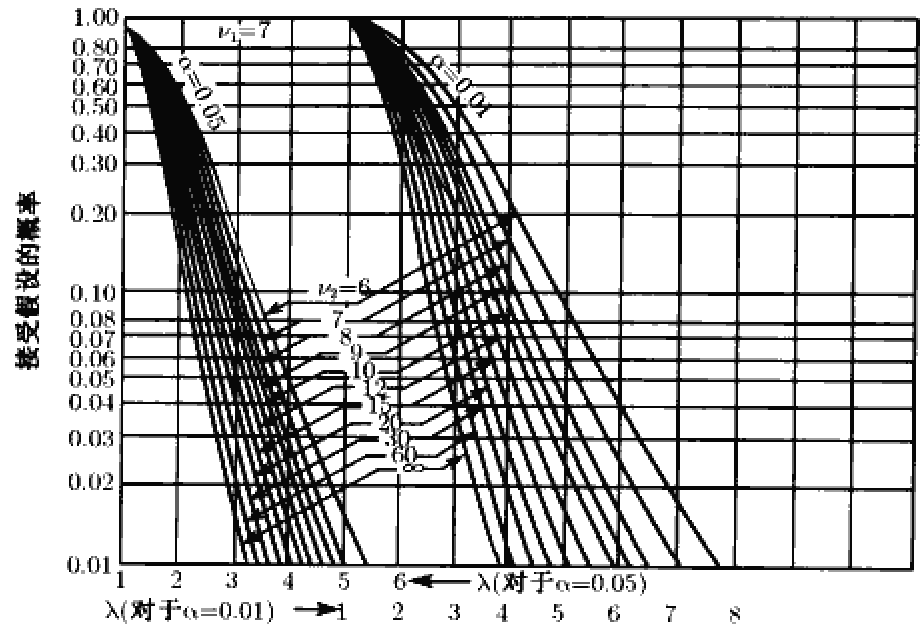
$\nu_1=4$
$\alpha=0.05$
$\alpha=0.01$
$\nu_2=6$, 7, 8, 9, 10, 12, 15, 20, 30, 60, ∞
接受假设的概率
1.00 0.80 0.70 0.60 0.50 0.40 0.30 0.20 0.10 0.08 0.07 0.06 0.05 0.04 0.03 0.02 0.01
1 2 3 4 5 6 7 8 ← λ(对于α=0.05)
λ(对于α=0.01)→1 2 3 4 5 6 7 8 9 10 11 12

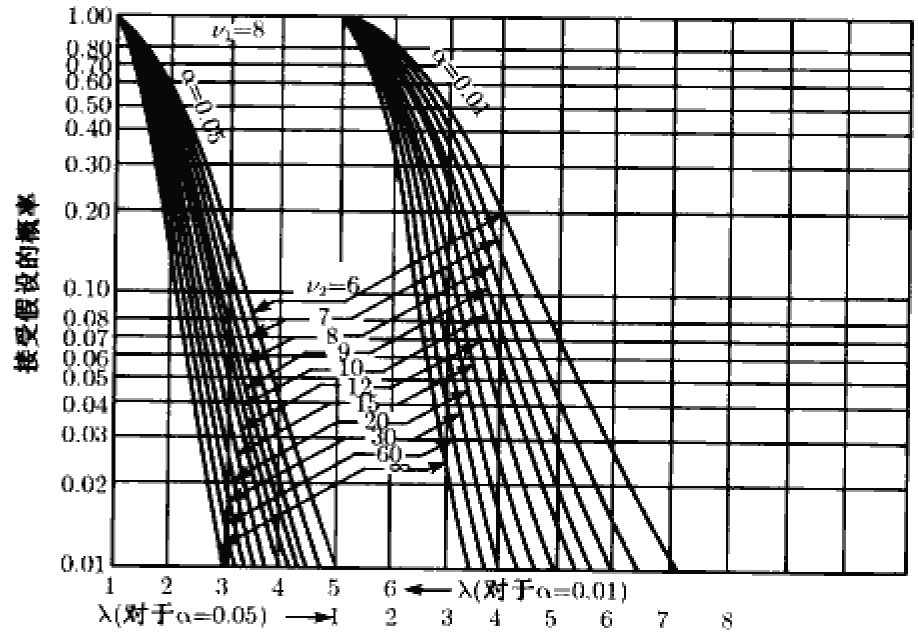


$\nu_1=5$
$\alpha=0.05$
$\alpha=0.01$
$\nu_2=6$, 7, 8, 9, 10, 12, 15, 20, 30, 60, ∞
接受假设的概率
1.00 0.80 0.70 0.60 0.50 0.40 0.30 0.20 0.10 0.08 0.07 0.06 0.05 0.04 0.03 0.02 0.01
1 2 3 4 5 6 7 ← λ(对于α=0.05)
λ(对于α=0.01) → 1 2 3 4 5 6 7 8 9 10



$\nu_1=6$
$\alpha=0.05$
$\alpha=0.01$
$\nu_2=6$, 7, 8, 9, 10, 12, 15, 20, 30, 60, ∞
接受假设的概率
1.00 0.80 0.70 0.60 0.50 0.40 0.30 0.20 0.10 0.08 0.07 0.06 0.05 0.04 0.03 0.02 0.01
1 2 3 4 5 6 7 ← λ(对于α=0.01)
λ(对于α=0.05) → 1 2 3 4 5 6 7 8 9


$\nu_1=7$
α=0.05
α=0.01
$\nu_2=6$
7
8
9
10
12
15
20
30
60
∞
接受假设的概率
1.00
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1 2 3 4 5 6 ← λ(对于α=0.05)
λ(对于α=0.01) → 1 2 3 4 5 6 7 8


$\nu_1=8$
α=0.05
α=0.01
$\nu_2=6$
7
8
9
10
12
15
20
30
60
∞
接受假设的概率
1.00
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1 2 3 4 5 6 ← λ(对于α=0.01)
λ(对于α=0.05) → 1 2 3 4 5 6 7 8

VII. 学生化极差统计量的百分位数①

$q_{0.01}(p, f)$

f \ p	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	90	135	164	186	202	216	227	237	246	253	260	266	272	272	282	286	290	294	298
2	14.0	19.0	22.3	24.7	26.6	28.2	29.5	30.7	31.7	32.6	33.4	31.4	34.8	35.4	36.0	36.5	37.0	37.5	37.9
3	8.26	10.6	12.2	13.3	14.2	15.0	15.6	16.2	16.7	17.1	17.5	17.9	18.2	18.5	18.8	19.1	19.3	19.5	19.8
4	6.51	8.12	9.17	9.96	10.6	11.1	11.5	11.9	12.3	12.6	12.8	13.1	13.3	13.5	13.7	13.9	14.1	14.2	14.4
5	5.70	6.97	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48	10.70	10.89	11.08	11.24	11.40	11.55	11.68	11.81	11.93
6	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30	9.49	9.65	9.81	9.95	10.08	10.21	10.32	10.43	10.54
7	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55	8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65
8	4.74	5.63	6.20	6.63	6.96	7.24	7.47	7.68	7.87	8.03	8.18	8.31	8.44	8.55	8.66	8.76	8.85	8.94	9.03
9	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.32	7.49	7.65	7.78	7.91	8.03	8.13	8.23	8.32	8.41	8.49	8.57
10	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36	7.48	7.60	7.71	7.81	7.91	7.99	8.07	8.15	8.22
11	4.39	5.14	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95
12	4.32	5.04	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7.73
13	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79	6.90	7.01	7.10	7.19	7.27	7.34	7.42	7.48	7.55
14	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66	6.77	6.87	6.96	7.05	7.12	7.20	7.27	7.33	7.39
15	4.17	4.83	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55	6.66	6.76	6.84	6.93	7.00	7.07	7.14	7.20	7.26
16	4.13	4.78	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
17	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38	6.48	6.57	6.66	6.73	6.80	6.87	6.94	7.00	7.05
18	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31	6.41	6.50	6.58	6.65	6.72	6.79	6.85	6.91	6.96
19	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
20	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19	6.29	6.37	6.45	6.52	6.59	6.65	6.71	6.76	6.82
24	3.96	4.54	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61
30	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
40	3.82	4.37	4.70	4.93	5.11	5.27	5.39	5.50	5.60	5.69	5.77	5.84	5.90	5.96	6.02	6.07	6.12	6.17	6.21
60	3.76	4.28	4.60	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60	5.67	5.73	5.79	5.84	5.89	5.93	5.98	6.02
120	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.38	5.44	5.51	5.56	5.61	5.66	5.71	5.75	5.79	5.83
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65

f = 自由度

① 摘自 J. M. May, "Extended and Corrected Tables of the Upper Percentage Points of the Studentized Range," *Biometrika*, Vol. 39, pp. 192–193, 1952. 承蒙*Biometrika*托管理事的惠允.

$q_{0.05}(p, f)$

(续)

f	p=2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	18.1	26.7	32.8	37.2	40.5	43.1	45.4	47.3	49.1	50.6	51.9	53.2	54.3	55.4	56.3	57.2	58.0	58.8	59.6
2	6.09	8.28	9.80	10.89	11.73	12.43	13.03	13.54	13.99	14.39	14.75	15.08	15.38	15.65	15.91	16.14	16.36	16.57	16.77
3	4.50	5.88	6.83	7.51	8.04	8.47	8.85	9.18	9.46	9.72	9.95	10.16	10.35	10.52	10.69	10.84	10.98	11.12	11.24
4	3.93	5.00	5.76	6.31	6.73	7.06	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.67	8.80	8.92	9.03	9.14	9.24
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	3.46	4.34	4.90	5.31	5.63	5.89	6.12	6.32	6.49	6.65	6.79	6.92	7.04	7.14	7.24	7.34	7.43	7.51	7.59
7	3.34	4.16	4.68	5.06	5.35	5.59	5.80	5.99	6.15	6.29	6.42	6.54	6.65	6.75	6.84	6.93	7.01	7.08	7.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	3.20	3.95	4.42	4.76	5.02	5.24	5.43	5.60	5.74	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.65
10	3.15	3.88	4.33	4.66	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.12	6.20	6.27	6.34	6.41	6.47
11	3.11	3.82	4.26	4.58	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98	6.06	6.14	6.20	6.27	6.33
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.40	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	3.06	3.73	4.15	4.46	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	6.00	6.06	6.11
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.56	5.64	5.72	5.79	5.86	5.92	5.98	6.03
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72	5.79	5.85	5.91	5.96
16	3.00	3.65	4.05	4.34	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	2.98	3.62	4.02	4.31	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.55	5.61	5.68	5.74	5.79	5.84
18	2.97	3.61	4.00	4.28	4.49	4.67	4.83	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	6.69	5.74	5.79
19	2.96	3.59	3.98	4.26	4.47	4.64	4.79	4.92	5.04	5.14	5.23	5.32	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	2.95	3.58	3.96	4.24	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.50	5.56	5.61	5.66	5.71
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.50	5.55	5.59
30	2.89	3.48	3.84	4.11	4.30	4.46	4.60	4.72	4.83	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.48
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.74	4.82	4.90	4.98	5.05	5.11	5.17	5.22	5.27	5.32	5.36
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	2.80	3.36	3.69	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
∞	2.77	3.32	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.84	4.98	4.93	4.97	5.01

VIII. 带有控制的比较处理的 Dunnett 检验的临界值①

$d_{0.05}(a-1, f)$

双侧比较

f	$a-1=$ 处理均值数 (不含控制)								
	1	2	3	4	5	6	7	8	9
5	2.57	3.03	3.29	3.48	3.62	3.73	3.82	3.90	3.97
6	2.45	2.86	3.10	3.26	3.39	3.49	3.57	3.64	3.71
7	2.36	2.75	2.97	3.12	3.24	3.33	3.41	3.47	3.53
8	2.31	2.67	2.88	3.02	3.13	3.22	3.29	3.35	3.41
9	2.26	2.61	2.81	2.95	3.05	3.14	3.20	3.26	3.32
10	2.23	2.57	2.76	2.89	2.99	3.07	3.14	3.19	3.24
11	2.20	2.53	2.72	2.84	2.94	3.02	3.08	3.14	3.19
12	2.18	2.50	2.68	2.81	2.90	2.98	3.04	3.09	3.14
13	2.16	2.48	2.65	2.78	2.87	2.94	3.00	3.06	3.10
14	2.14	2.46	2.63	2.75	2.84	2.91	2.97	3.02	3.07
15	2.13	2.44	2.61	2.73	2.82	2.89	2.95	3.00	3.04
16	2.12	2.42	2.59	2.71	2.80	2.87	2.92	2.97	3.02
17	2.11	2.41	2.58	2.69	2.78	2.85	2.90	2.95	3.00
18	2.10	2.40	2.56	2.68	2.76	2.83	2.89	2.94	2.98
19	2.09	2.39	2.55	2.66	2.75	2.81	2.87	2.92	2.96
20	2.09	2.38	2.54	2.65	2.73	2.80	2.86	2.90	2.95
24	2.06	2.35	2.51	2.61	2.70	2.76	2.81	2.86	2.90
30	2.04	2.32	2.47	2.58	2.66	2.72	2.77	2.82	2.86
40	2.02	2.29	2.44	2.54	2.62	2.68	2.73	2.77	2.81
60	2.00	2.27	2.41	2.51	2.58	2.64	2.69	2.73	2.77
120	1.98	2.24	2.38	2.47	2.55	2.60	2.65	2.69	2.73
∞	1.96	2.21	2.35	2.44	2.51	2.57	2.61	2.65	2.69

$d_{0.01}(a-1, f)$

双侧比较

f	$a-1=$ 处理均值数 (不含控制)								
	1	2	3	4	5	6	7	8	9
5	4.03	4.63	4.98	5.22	5.41	5.56	5.69	5.80	5.89
6	3.71	4.21	4.51	4.71	4.87	5.00	5.10	5.20	5.28
7	3.50	3.95	4.21	4.39	4.53	4.64	4.74	4.82	4.89
8	3.36	3.77	4.00	4.17	4.29	4.40	4.48	4.56	4.62
9	3.25	3.63	3.85	4.01	4.12	4.22	4.30	4.37	4.43
10	3.17	3.53	3.74	3.88	3.99	4.08	4.16	4.22	4.28
11	3.11	3.45	3.65	3.79	3.89	3.98	4.05	4.11	4.16
12	3.05	3.39	3.58	3.71	3.81	3.89	3.96	4.02	4.07
13	3.01	3.33	3.52	3.65	3.74	3.82	3.89	3.94	3.99
14	2.98	3.29	3.47	3.59	3.69	3.76	3.83	3.88	3.93
15	2.95	3.25	3.43	3.55	3.64	3.71	3.78	3.83	3.88
16	2.92	3.22	3.39	3.51	3.60	3.76	3.73	3.78	3.83
17	2.90	3.19	3.36	3.47	3.56	3.63	3.69	3.74	3.79
18	2.88	3.17	3.33	3.44	3.53	3.60	3.66	3.71	3.75
19	2.86	3.15	3.31	3.42	3.50	3.57	3.63	3.68	3.72
20	2.85	3.13	3.29	3.40	3.48	3.55	3.60	3.65	3.69
24	2.80	3.07	3.22	3.32	3.40	3.47	3.52	3.57	3.61
30	2.75	3.01	3.15	3.25	3.33	3.39	3.44	3.49	3.52
40	2.70	2.95	3.09	3.19	3.26	3.32	3.37	3.41	3.44
60	2.66	2.90	3.03	3.12	3.19	3.25	3.29	3.33	3.37
120	2.62	2.85	2.97	3.06	3.12	3.18	3.22	3.26	3.29
∞	2.58	2.79	2.92	3.00	3.06	3.11	3.15	3.19	3.22

$f=$ 自由度

① 承蒙惠允，摘自 C. W. Dunnett, "New Tables for Multiple Comparison with a Control," *Biometrics*, Vol. 20, No. 3, 1964, 以及 C. W. Dunnett, "A Multiple Comparison Procedure for Comparing Several Treatments with a Control," *Journal of the American Statistical Association*, Vol. 50, 1955.

$d_{0.05}(a-1, f)$

单侧比较 (续)

f	$a-1=$ 处理均值数 (不含控制)								
	1	2	3	4	5	6	7	8	9
5	2.02	2.44	2.68	2.85	2.98	3.08	3.16	3.24	3.30
6	1.94	2.34	2.56	2.71	2.83	2.92	3.00	3.07	3.12
7	1.89	2.27	2.48	2.62	2.73	2.82	2.89	2.95	3.01
8	1.86	2.22	2.42	2.55	2.66	2.74	2.81	2.87	2.92
9	1.83	2.18	2.37	2.50	2.60	2.68	2.75	2.81	2.86
10	1.81	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81
11	1.80	2.13	2.31	2.44	2.53	2.60	2.67	2.72	2.77
12	1.78	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74
13	1.77	2.09	2.27	2.39	2.48	2.55	2.61	2.66	2.71
14	1.76	2.08	2.25	2.37	2.46	2.53	2.59	2.64	2.69
15	1.75	2.07	2.24	2.36	2.44	2.51	2.57	2.62	2.67
16	1.75	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65
17	1.74	2.05	2.22	2.33	2.42	2.49	2.54	2.59	2.64
18	1.73	2.04	2.21	2.32	2.41	2.48	2.53	2.58	2.62
19	1.73	2.03	2.20	2.31	2.40	2.47	2.52	2.57	2.61
20	1.72	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60
24	1.71	2.01	2.17	2.28	2.36	2.43	2.48	2.53	2.57
30	1.70	1.99	2.15	2.25	2.33	2.40	2.45	2.50	2.54
40	1.68	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51
60	1.67	1.95	2.10	2.21	2.28	2.35	2.39	2.44	2.48
120	1.66	1.93	2.08	2.18	2.26	2.32	2.37	2.41	2.45
∞	1.64	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42

$d_{0.01}(a-1, f)$

单侧比较

f	$a-1=$ 处理均值数 (不含控制)								
	1	2	3	4	5	6	7	8	9
5	3.37	3.90	4.21	4.43	4.60	4.73	4.85	4.94	5.03
6	3.14	3.61	3.88	4.07	4.21	4.33	4.43	4.51	4.59
7	3.00	3.42	3.66	3.83	3.96	4.07	4.15	4.23	4.30
8	2.90	3.29	3.51	3.67	3.79	3.88	3.96	4.03	4.09
9	2.82	3.19	3.40	3.55	3.66	3.75	3.82	3.89	3.94
10	2.76	3.11	3.31	3.45	3.56	3.64	3.71	3.78	3.83
11	2.72	3.06	3.25	3.38	3.48	3.56	3.63	3.69	3.74
12	2.68	3.01	3.19	3.32	3.42	3.50	3.56	3.62	3.67
13	2.65	2.97	3.15	3.27	3.37	3.44	3.51	3.56	3.61
14	2.62	2.94	3.11	3.23	3.32	3.40	3.46	3.51	3.56
15	2.60	2.91	3.08	3.20	3.29	3.36	3.42	3.47	3.52
16	2.58	2.88	3.05	3.17	3.26	3.33	3.39	3.44	3.48
17	2.57	2.86	3.03	3.14	3.23	3.30	3.36	3.41	3.45
18	2.55	2.84	3.01	3.12	3.21	3.27	3.33	3.38	3.42
19	2.54	2.83	2.99	3.10	3.18	3.25	3.31	3.36	3.40
20	2.53	2.81	2.97	3.08	3.17	3.23	3.29	3.34	3.38
24	2.49	2.77	2.92	3.03	3.11	3.17	3.22	3.27	3.31
30	2.46	2.72	2.87	2.97	3.05	3.11	3.16	3.21	3.24
40	2.42	2.68	2.82	2.92	2.99	3.05	3.10	3.14	3.18
60	2.39	2.64	2.78	2.87	2.94	3.00	3.04	3.08	3.12
120	2.36	2.60	2.73	2.82	2.89	2.94	2.99	3.03	3.06
∞	2.33	2.56	2.68	2.77	2.84	2.89	2.93	2.97	3.00

IX. 正交多项式的系数①

	X_j	$n=3$		$n=4$			$n=5$				$n=6$					$n=7$					
		P_1	P_2	P_1	P_2	P_3	P_1	P_2	P_3	P_4	P_1	P_2	P_3	P_4	P_5	P_1	P_2	P_3	P_4	P_5	P_6
	1	−1	1	−3	1	−1	−2	2	−1	1	−5	5	−5	1	−1	−3	5	−1	3	−1	1
	2	0	−2	−1	−1	3	−1	−1	2	−4	−3	−1	7	−3	5	−2	0	1	−7	4	−6
	3	1	1	1	−1	−3	0	−2	0	6	−1	−4	4	2	−10	−1	−3	1	1	−5	15
	4			3	1	1	1	−1	−2	−4	1	−4	−4	2	10	0	−4	0	6	0	−20
	5						2	2	1	1	3	−1	−7	−3	−5	1	−3	−1	1	5	15
	6										5	5	5	1	1	2	0	−1	−7	−4	−6
	7															3	5	1	3	1	1
$\sum_{j=1}^{n}\{P_i(X_j)\}^2$		2	6	20	4	20	10	14	10	70	70	84	180	28	252	28	84	6	154	84	924
	λ	1	3	2	1	$\frac{10}{3}$	1	1	$\frac{5}{6}$	$\frac{35}{12}$	2	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{7}{12}$	$\frac{21}{10}$	1	1	$\frac{1}{6}$	$\frac{7}{12}$	$\frac{7}{20}$	$\frac{77}{60}$

	X_j	$n=8$						$n=9$						$n=10$					
		P_1	P_2	P_3	P_4	P_5	P_6	P_1	P_2	P_3	P_4	P_5	P_6	P_1	P_2	P_3	P_4	P_5	P_6
	1	−7	7	−7	7	−7	1	−4	28	−14	14	−4	4	−9	6	−42	18	−6	3
	2	−5	1	5	−13	23	−5	−3	7	7	−21	11	−17	−7	2	14	−22	14	−11
	3	−3	−3	7	−3	−17	9	−2	−8	13	−11	−4	22	−5	−1	35	−17	−1	10
	4	−1	−5	3	9	−15	−5	−1	−17	9	9	−9	1	−3	−3	31	3	−11	6
	5	1	−5	−3	9	15	−5	0	−20	0	18	0	−20	−1	−4	12	18	−6	−8
	6	3	−3	−7	−3	17	9	1	−17	−9	9	9	1	1	−4	−12	18	6	−8
	7	5	1	−5	−13	−23	−5	2	−8	−13	−11	4	22	3	−3	−31	3	11	6
	8	7	7	7	7	7	1	3	7	−7	−21	−11	−17	5	−1	−35	−17	1	10
	9							4	28	14	14	4	4	7	2	−14	−22	−14	−11
	10													9	6	42	18	6	3
$\sum_{j=1}^{n}\{P_i(X_j)\}^2$		168	168	264	616	2 184	264	60	2 772	990	2 002	468	1 980	330	132	8 580	2 860	780	660
	λ	2	1	$\frac{2}{3}$	$\frac{7}{12}$	$\frac{7}{10}$	$\frac{11}{60}$	1	3	$\frac{5}{6}$	$\frac{7}{12}$	$\frac{3}{20}$	$\frac{11}{60}$	2	$\frac{1}{2}$	$\frac{5}{3}$	$\frac{5}{12}$	$\frac{1}{10}$	$\frac{11}{240}$

① 承蒙惠允，摘自 E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Vol. 1, 3rd edition, Cambridge University Press, Cambridge, 1966.

X. 2^{k-p} 分式析因设计的别名关系 ($k\leqslant 15, n\leqslant 64$)

3 个因子的设计

(a) 2^{3-1} 分式析因设计　　分辨度 III

设计生成元

$C = AB$

定义关系: $I = ABC$

别名

$A = BC$

$B = AC$

$C = AB$

4 个因子的设计

(b) 2^{4-1} 分式析因设计　　分辨度 IV

设计生成元

$D = ABC$

定义关系: $I = ABCD$

别名

$A = BCD$

$B = ACD$

$C = ABD$

$D = ABC$

$AB = CD$

$AC = BD$

$AD = BC$

5 个因子的设计

(c) 2^{5-2} 分式析因设计　　分辨度 III

设计生成元

$D = AB \quad E = AC$

定义关系: $I = ABD = ACE = BCDE$

别名

$A = BD = CE$

$B = AD = CDE$

$C = AE = BDE$

$D = AB = BCE$

$E = AC = BCD$

$BC = DE = ACD = ABE$

$CD = BE = ABC = ADE$

(d) 2^{5-1} 分式析因设计　　分辨度 V

设计生成元

$E = ABCD$

定义关系: $I = ABCDE$

别名

每一个主效应都与一个四因子交互作用别名

$AB = CDE$	$BD = ACE$
$AC = BDE$	$BE = ACD$
$AD = BCE$	$CD = ABE$
$AE = BCD$	$CE = ABD$
$BC = ADE$	$DE = ABC$

8 个区组中的 2 个: $AB = CDE$

(续)

6 个因子的设计

(e) 2^{6-3} 分式析因设计 分辨度 III

设计生成元

$D = AB \quad E = AC \quad F = BC$

定义关系: $I = ABD = ACE = BCDE = BCF = ACDF = ABEF = DEF$

别名

$A=BD=CE=CDF=BEF$	$E=AC=DF=BCD=ABF$
$B=AD=CF=CDE=AEF$	$F=BC=DE=ACD=ABE$
$C=AE=BF=BDE=ADF$	$CD=BE=AF=ABC=ADE=BDF=CEF$
$D=AB=EF=BCE=ACF$	

(f) 2^{6-2} 分式析因设计 分辨度 IV

设计生成元

$E = ABC \quad F = BCD$

定义关系: $I = ABCE = BCDF = ADEF$

别名

$A = BCE = DEF$	$AB = CE$
$B = ACE = CDF$	$AC = BE$
$C = ABE = BDF$	$AD = EF$
$D = BCF = AEF$	$AE = BC = DF$
$E = ABC = ADF$	$AF = DE$
$F = BCD = ADE$	$BD = CF$
$ABD = CDE = ACF = BEF$	$BF = CD$
$ACD = BDE = ABF = CEF$	

8 个区组中的 2 个: $ABD = CDE = ACF = BEF$

(g) 2^{6-1} 分式析因设计 分辨度 VI

设计生成元

$F = ABCDE$

定义关系: $I = ABCDEF$

别名

每个主效应都与一个五因子交互作用别名

每个二因子交互作用都与一个四因子交互作用别名

$ABC = DEF$	$ACE = BDF$
$ABD = CEF$	$ACF = BDE$
$ABE = CDF$	$ADE = BCF$
$ABF = CDE$	$ADF = BCE$
$ACD = BEF$	$AEF = BCD$

16 个区组中的 2 个: $ABC = DEF$

8 个区组中的 4 个: $AB = CDEF$, $ACD = BEF$, $AEF = BCD$

（续）

7 个因子的设计

(h) 2^{7-4} 分式析因设计　　分辨度 III

设计生成元

$D=AB \quad E=AC \quad F=BC \quad G=ABC$

定义关系: $I=ABD=ACE=BCDE=BCF$
$=ACDF=ABEF=DEF=ABCG$
$=CDG=BEG=ADEG=AFG$
$=BDFG=CEFG=ABCDEFG$

别名

$A=BD=CE=FG \quad E=AC=DF=BG$
$B=AD=CF=EG \quad F=BC=DE=AG$
$C=AE=BF=DG \quad G=CD=BE=AF$
$D=AB=EF=CG$

(i) 2^{7-3} 分式析因设计　　分辨度 IV

设计生成元

$E=ABC \quad F=BCD \quad G=ACD$

定义关系: $I=ABCE=BCDF=ADEF=ACDG=BDEG=ABFG=CEFG$

别名

$A=BCE=DEF=CDG=BFG \quad AB=CE=FG \quad E=ABC=ADF=BDG=CFG \quad AF=DE=BG$
$B=ACE=CDF=DEG=AFG \quad AC=BE=DG \quad F=BCD=ADE=ABG=CEG \quad AG=CD=BF$
$C=ABE=BDF=ADG=EFG \quad AD=EF=CG \quad G=ACD=BDE=ABF=CEF \quad BD=CF=EG$
$D=BCF=AEF=ACG=BEG \quad AE=BC=DF$

$ABD=CDE=ACF=BEF=BCG=AEG=DFG$

8 个区组中的 2 个: $ABD=CDE=ACF=BEF=BCG=AEG=DFG$

(j) 2^{7-2} 分式析因设计　　分辨度 IV

设计生成元

$F=ABCD \quad G=ABDE$

定义关系: $I=ABCDF=ABDEG=CEFG$

别名

$A=$	$AB=CDF=DEG$	$BC=ADF$	$CE=FG$	$ACE=AFG$
$B=$	$AC=BDF$	$BD=ACF=AEG$	$CF=ABD=EG$	$ACG=AEF$
$C=EFG$	$AD=BCF=BEG$	$BE=ADG$	$CG=EF$	$BCE=BFG$
$D=$	$AE=BDG$	$BF=ACD$	$DE=ABG$	$BCG=BEF$
$E=CFG$	$AF=BCD$	$BG=ADE$	$DF=ABC$	$CDE=DFG$
$F=CEG$	$AG=BDE$	$CD=ABF$	$DG=ABE$	$CDG=DEF$
$G=CEF$				

16 个区组中的 2 个: $ACE=AFG$　　8 个区组中的 4 个 : $ACE=AFG$
$BCE=BFG$
$AB=CDF=DEG$

(k) 2^{7-1} 分式析因设计　　分辨度 VII

设计生成元

$G=ABCDEF$

定义关系: $I=ABCDEFG$

别名

每一个主效应都与一个六因子交互作用别名
每一个二因子交互作用都与一个五因子交互作用别名
每一个三因子交互作用都与一个四因子交互作用别名

32 个区组中的 2 个: ABC　　16 个区组中的 4 个: ABC
CEF
CDG

(续)

8 个因子的设计

(l) 2^{8-4} 分式析因设计 分辨度 IV

设计生成元

$E = BCD \quad F = ACD \quad G = ABC \quad H = ABD$

定义关系: $I = BCDE = ACDF = ABEF = ABCG = ADEG = BDFG$

$= CEFG = ABDH = ACEH = BCFH = DEFH$

$= CDGH = BEGH = AFGH = ABCDEFGH$

别名

$A = CDF = BEF = BCG = DEG = BDH = CEH = FGH$	$AB = EF = CG = DH$
$B = CDE = AEF = ACG = DFG = ADH = CFH = EGH$	$AC = DF = BG = EH$
$C = BDE = ADF = ABG = EFG = AEH = BFH = DGH$	$AD = CF = EG = BH$
$D = BCE = ACF = AEG = BFG = ABH = EFH = CGH$	$AE = BF = DG = CH$
$E = BCD = ABF = ADG = CFG = ACH = DFH = BGH$	$AF = CD = BE = GH$
$F = ACD = ABE = BDG = CEG = BCH = DEH = AGH$	$AG = BC = DE = FH$
$G = ABC = ADE = BDF = CEF = CDH = BEH = AFH$	$AH = BD = CE = FG$
$H = ABD = ACE = BCF = DEF = CDG = BEG = AFG$	

8 个区组中的 2 个: $AB = EF = CG = DH$

(m) 2^{8-3} 分式析因设计 分辨度 IV

设计生成元

$F = ABC \quad G = ABD \quad H = BCDE$

定义关系: $I = ABCF = ABDG = CDFG = BCDEH = ADEFH = ACEGH = BEFGH$

别名

$A = BCF = BDG$	$AE = DFH = CGH$	$DE = BCH = AFH$
$B = ACF = ADG$	$AF = BC = DEH$	$DH = BCE = AEF$
$C = ABF = DFG$	$AG = BD = CEH$	$EF = ADH = BGH$
$D = ABG = CFG$	$AH = DEF = CEG$	$EG = ACH = BFH$
$E =$	$BE = CDH = FGH$	$EH = BCD = ADF = ACG = BFG$
$F = ABC = CDG$	$BH = CDE = EFG$	$FH = ADE = BEG$
$G = ABD = CDF$	$CD = FG = BEH$	$GH = ACE = BEF$
$H =$	$CE = BDH = AGH$	$ABE = CEF = DEG$
$AB = CF = DG$	$CG = DF = AEH$	$ABH = CFH = DGH$
$AC = BF = EGH$	$CH = BDE = AEG$	$ACD = BDF = BCG = AFG$
$AD = BG = EFH$		

16个区组中的 2 个 : $ABE=CEF=DEG$ 8个区组中的4个 : $ABE=CEF=DEG$

$ABH=CFH=DGH$

$EH=BCD=ADF=ACG=BFG$

（续）

8 个因子的设计（续）

(n) 2^{8-2} 分式析因设计　　分辨度 V

设计生成元

$G = ABCD \quad H = ABEF$

定义关系: $I = ABCDG = ABEFH = CDEFGH$

别名

$AB = CDG = EFH$	$BG = ACD$	$EF = ABH$	$ADH =$	$BFG =$
$AC = BDG$	$BH = AEF$	$EG =$	$AEG =$	$BGH =$
$AD = BCG$	$CD = ABG$	$EH = ABF$	$AFG =$	$CDE = FGH$
$AE = BFH$	$CE =$	$FG =$	$AGH =$	$CDF = EGH$
$AF = BEH$	$CF =$	$FH = ABE$	$BCE =$	$CDH = EFG$
$AG = BCD$	$CG = ABD$	$GH =$	$BCF =$	$CEF = DGH$
$AH = BEF$	$CH =$	$ACE =$	$BCH =$	$CEG = DFH$
$BC = ADG$	$DE =$	$ACF =$	$BDE =$	$CEH = DFG$
$BD = ACG$	$DF =$	$ACH =$	$BDF =$	$CFG = DEH$
$BE = AFH$	$DG = ABC$	$ADE =$	$BDH =$	$CFH = DEG$
$BF = AEH$	$DH =$	$ADF =$	$BEG =$	$CGH = DEF$

32 个区组中的 2 个: $CDE = FGH$　16 个区组中的 4 个: $CDE = FGH$
ACF
BDH

9 个因子的设计

(o) 2^{9-5} 分式析因设计　　分辨度 III

设计生成元

$E = ABC \quad F = BCD \quad G = ACD \quad H = ABD \quad J = ABCD$

定义关系 : $I = ABCE = BCDF = ADEF = ACDG = BDEG = ABFG = CEFG = ABDH$
$= CDEH = ACFH = BEFH = BCGH = AEGH = DFGH = ABCDEFGH$
$= ABCDJ = DEJ = AFJ = BCEFJ = BGJ = ACEGJ = CDFGJ$
$= ABDEFGJ = CHJ = ABEHJ = BDFHJ = ABCDEFHJ$
$= ADGHJ = BCDEFGHJ = ABCFGHJ = EFGHJ$

别名

$A = FJ$
$B = GJ$
$C = HJ$
$D = EJ$
$E = DJ$
$F = AJ$
$G = BJ$
$H = CJ$
$J = DE = AF = BG = CH$
$AB = CE = FG = DH$
$AC = BE = DG = FH$
$AD = EF = CG = BH$
$AE = BC = DF = GH$
$AG = CD = BF = EH$
$AH = BD = CF = EG$

8 个区组中的 2 个: $AB = CE = FG = DH$

(续)

9 个因子的设计 (续)

(p) 2^{9-4} 分式析因设计 分辨度 IV

设计生成元

$F = BCDE \quad G = ACDE \quad H = ABDE \quad J = ABCE$

定义关系：$I = BCDEF = ACDEG = ABFG = ABDEH = ACFH = BCGH$
$= DEFGH = ABCEJ = ADFJ = BDGJ = CEFGJ = CDHJ$
$= BEFHJ = AEGHJ = ABCDFGHJ$

别名

$A=BFG=CFH=DFJ$	$AD=CEG=BEH=FJ$	$BJ=ACE=DG=EFH$
$B=AFG=CGH=DGJ$	$AE=CDG=BDH=BCJ=GHJ$	$CD=BEF=AEG=HJ$
$C=AFH=BGH=DHJ$	$AF=BG=CH=DJ$	$CE=BDF=ADG=ABJ=FGJ$
$D=AFJ=BGJ=CHJ$	$AG=CDE=BF=EHJ$	$CJ=ABE=EFG=DH$
$E=$	$AH=BDE=CF=EGJ$	$DE=BCF=ACG=ABH=FGH$
$F=ABG=ACH=ADJ$	$AJ=BCE=DF=EGH$	$EF=BCD=DGH=CGJ=BHJ$
$G=ABF=BCH=BDJ$	$BC=DEF=GH=AEJ$	$EG=ACD=DFH=CFJ=AHJ$
$H=ACF=BCG=CDJ$	$BD=CEF=AEH=GJ$	$EH=ABD=DFG=BFJ=AGJ$
$J=ADF=BDG=CDH$	$BE=CDF=ADH=ACJ=FHJ$	$EJ=ABC=CFG=BFH=AGH$
$AB=FG=DEH=CEJ$	$BH=ADE=CG=EFJ$	$AEF=BEG=CEH=DEJ$
$AC=DEG=FH=BEJ$		

16个区组中的2个: $AEF=BEG=CEH=DEJ$ 8 个区组中的4个: $AEF = BEG = CEH = DEJ$
$AB = FG = DEH = CEJ$
$CD = BEF = AEG = HJ$

(q) 2^{9-3} 分式析因设计 分辨度 IV

设计生成元

$G = ABCD \quad H = ACEF \quad J = CDEF$

定义关系: $I = ABCDG = ACEFH = BDEFGH = CDEFJ = ABEFGJ = ADHJ = BCGHJ$

别名

$A = DHJ$	$AC = BDG = EFH$	$BF =$
$B =$	$AD = BCG = HJ$	$BG = ACD = CHJ$
$C =$	$AE = CFH$	$BH = CGJ$
$D = AHJ$	$AF = CEH$	$BJ = CGH$
$E =$	$AG = BCD$	$CD = ABG = EFJ$
$F =$	$AH = CEF = DJ$	$CE = AFH = DFJ$
$G =$	$AJ = DH$	$CF = AEH = DEJ$
$H = ADJ$	$BC = ADG = GHJ$	$CG = ABD = BHJ$
$J = ADH$	$BD = ACG$	$CH = AEF = BGJ$
$AB = CDG$	$BE =$	$CJ = DEF = BGH$
$DE = CFJ$	$GJ = BCH$	$AFJ = BEG = DFH$
$DF = CEJ$	$ABE = EGJ$	$AGH = DGJ$
$DG = ABC$	$ABF = EGJ$	$AGJ = BEF = DGH$
$EF = ACH = CDJ$	$ABH = BDJ$	$BCE =$
$EG =$	$ABJ = EFG = BDH$	$BCF =$
$EH = ACF$	$ACJ = CDH$	$BDE = FGH$
$EJ = CDF$	$ADE = EHJ$	$BDF = EGH$
$FG =$	$ADF = FHJ$	$BEH = DFG$
$FH = ACE$	$AEG = BFJ$	$BFH = BEG$
$FJ = CDE$	$AEJ = BFG = DEH$	$CEG =$
$GH = BCJ$	$AFG = BEJ$	$CFG =$

32 个区组中的 2 个: CFG 16 个区组中的 4 个: $CFG =$
$AGJ = BEF = DGH$
$ADE = EHJ$

（续）

10 个因子的设计

(r) 2^{10-6} 分式析因设计　　　　分辨度 III

设计生成元

$E = ABC \quad F = BCD \quad G = ACD \quad H = ABD \quad J = ABCD \quad K = AB$

定义关系：$I = ABCE = BCDF = ADEF = ACDG = BDEG = ABFG = CEFG = ABDH$
$= CDEH = ACFH = BEFH = BCGH = AEGH = DFGH = ABCDEFGH$
$= ABCDJ = DEJ = AFJ = BCEFJ = BGJ = ACEGJ = CDFGJ$
$= ABDEFGJ = CHJ = ABEHJ = BDFHJ = ACDEFHJ = ADGHJ$
$= BCDEGHJ = ABCFGHJ = EFGHJ = ABK = CEK = ACDFK$
$= BDEFK = BCDGK = ADEGK = FGK = ABCEFGK = DHK$
$= ABCDEHK = BCFHK = AEFHK = ACGHK = BEGHK = ABDFGHK$
$= CDEFGHK = CDJK = ABDEJK = BFJK = ACEFJK = AGJK$
$= BCEGJK = ABCDFGJK = DEFGJK = ABCHJK = EHJK = ADFHJK$
$= BCDEFHJK = BDGHJK = ACDEGHJK = CFGHJK = ABEFGHJK$

别名

$A = FJ = BK \qquad J = DE = AF = BG = CH$
$B = GJ = AK \qquad K = AB = CE = FG = DH$
$C = HJ = EK \qquad AC = BE = DG = FH$
$D = EJ = HK \qquad AD = EF = CG = BH$
$E = DJ = CK \qquad AE = BC = DF = GH$
$F = AJ = GK \qquad AG = CD = BF = EH = JK$
$G = BJ = FK \qquad AH = BD = CF = EG$
$H = CJ = DK$

8 个区组中的 2 个: $AG = CD = BF = EH = JK$

(s) 2^{10-5} 分式析因设计　　　　分辨度 IV

设计生成元

$F = ABCD \quad G = ABCE \quad H = ABDE \quad J = ACDE \quad K = BCDE$

定义关系：$I = ABCDF = ABCEG = DEFG = ABDEH = CEFH = CDGH = ABFGH$
$= ACDEJ = BEFJ = BDGJ = ACFGJ = BCHJ = ADFHJ = AEGHJ$
$= BCDEFGHJ = BCDEK = AEFK = ADGK = BCFGK = ACHK$
$= BDFHK = BEGHK = ACDEFGHK = ABJK = CDFJK = CEGJK$
$= ABDEFGJK = BEHJK = ABCEFHJK = ABCDGHJK = FGHJK$

别名

$A = EFK = DGK = CHK = BJK$
$B = EFJ = DGJ = CHJ = AJK$
$C = EFH = DGH = BHJ = AHK$
$D = EFG = CGH = BGJ = AGK$
$E = DFG = CFH = BFJ = AFK$
$F = DEG = CEH = BEJ = AEK$
$G = DEF = CDH = BDJ = ADK$
$H = CEF = CDG = BCJ = ACK$
$J = BEF = BDG = BCG = ABK$
$K = AEF = ADG = ACH = ABJ$
$AB = CDF = CEG = DEH = FGH = JK$
$AC = BDF = BEG = DEJ = FGJ = HK$
$AD = BCF = BEH = CEJ = FHJ = GK$
$AE = BCG = BDH = CDJ = GHJ = FK$
$AF = BCD = BGH = CGJ = DHJ = EK$
$AC = BCE = BFH = CFJ = EHJ = DK$
$AH = BDE = BFG = DFJ = EGJ = CK$
$AJ = CDE = CFG = DFH = EGH = BK$
$AK = EF = DG = CH = BJ$
$BC = ADF = AEG = HJ = DEK = FGK$
$BD = ACF = AEH = GJ = CEK = FHK$
$BE = ACG = ADH = FJ = CDK = GHK$
$BF = ACD = AGH = EJ = CGK = DHK$
$BG = ACE = AFH = DJ = CFK = EHK$
$BH = ADE = AFG = CJ = DFK = EGK$
$CD = ABF = GH = AEJ = BEK = FJK$
$CE = ABG = FH = ADJ = BDK = GJK$
$CF = ABD = EH = AGJ = BGK = DJK$
$CG = ABE = BH = AFJ = BFK = EJK$
$DE = FG = ABH = ACJ = BCK = HJK$
$DF = ABC = EG = AHJ = BHK = CJK$

16 个区组中的 2 个：$AK = EF = DG = CH = BJ$

8 个区组中的 4 个：$AK = EF = DG = CH = BJ$
$AJ = CDE = CFG = DFH = EGH = BK$
$AB = CDF = CEG = DEH = FGH = JK$

(续)

10 个因子的设计 (续)

(t) 2^{10-4} 分式析因设计 分辨度 IV

设计生成元

$G=BCDF \quad H=ACDF \quad J=ABDE \quad K=ABCE$

定义关系：$I=BCDFG=ACDFH=ABGH=ABDEJ=ACEFGJ=BCEFHJ$
$=DEGHJ=ABCEK=ADEFGK=BDEFHK=CEGHK$
$=CDJK=BFGJK=AFHJK=ABCDGHJK$

别名

$A=BGH$	$AD=CFH=BEJ$	$BK=ACE=FGJ$
$B=AGH$	$AE=BDJ=BCK$	$CD=BFG=AFH=JK$
$C=DJK$	$AF=CDH=HJK$	$CE=ABK=GHK$
$D=CJK$	$AG=BH$	$CF=BDG=ADH$
$E=$	$AH=CDF=BG=FJK$	$CG=BDF=EHK$
$F=$	$AJ=BDE=FHK$	$CH=ADF=EGK$
$G=ABH$	$AK=BCE=FHJ$	$CJ=DK$
$H=ABG$	$BC=DFG=AEK$	$CK=ABE=EGH=DJ$
$J=CDK$	$BD=CFG=AEJ$	$DE=ABJ=GHJ$
$K=CDJ$	$BE=ADJ=ACK$	$DF=BCG=ACH$
$AB=GH=DEJ=CEK$	$BF=CDG=GJK$	$DG=BCF=EHJ$
$AC=DFH=BEX$	$BJ=ADE=FGK$	$DH=ACF=EGJ$
$EF=$	$GJ=DEH=BFK$	$AEG=BEH=CFJ=DFK$
$EG=DHJ=CHK$	$GK=CEH=BFJ$	$AEH=BEG$
$EH=DGJ=CGK$	$HJ=DEG=AFK$	$AFG=BFH=CEJ=DEK$
$EJ=ABD=DGH$	$HK=CEG=AFJ$	$AGJ=CEF=BHJ$
$EK=ABC=CGH$	$ABF=FGH$	$AGK=DEF=BHK$
$FG=BCD=BJK$	$ACG=BCH=EFJ$	$BCJ=EFH=BDK$
$FH=ACD=AJK$	$ACJ=EFG=ADK$	$BEF=CHJ=DHK$
$FJ=BGK=AHK$	$ADG=BDH=EFK$	$CDE=EJK$
$FK=BGJ=AHJ$	$AEF=CGJ=DGK$	$CFK=DFJ$

32 个区组中的 2 个：$AGJ=CEF=BHJ$

16 个区组中的 4 个：$AGJ=CEF=BHJ$
$AGK=DEF=BHK$
$CD=BFG=AFH=JK$

11 个因子的设计

(u) 2^{11-7} 分式析因设计 分辨度 III

设计生成元

$E=ABC \quad F=BCD \quad G=ACD \quad H=ABD \quad J=ABCD \quad K=AB \quad L=AC$

定义关系：$I=ABCE=BCDF=ADEF=ACDG=BDEG=ABFG=CEFG=ABDH$
$=CDEH=ACFH=BEFH=BCGH=AEGH=DFGH=ABCDEFGH$
$=ABCDJ=DEJ=AFJ=BCEFJ=BGJ=ACEGJ=CDFGJ$

(续)

$= ABDEFGJ = CHJ = ABEHJ = BDFHJ = ACDEFHJ = ADGHJ$
$= BCDEGHJ = ABCFGHJ = EFGHJ = ABK = CEK = ACDFK$
$= BDEFK = BCDGK = ADEGK = FGK = ABCEFGK = DHK$
$= ABCDEHK = BCFHK = AEFHK = ACGHK = BEGHK$
$= ABDFGHK = CDEFGHK = CDJK = ABDEJK = BFJK = ACEFJK$
$= AGJK = BCEGJK = ABCDFGJK = DEFGJK = ABCHJK = EHJK$
$= ADFHJK = BCDEFHJK = BDGHJK = ACDEGHJK = CFGHJK$
$= ABEFGHJK = ACL = BEL = ABDFL = CDEFL = DGL = ABCDEGL$
$= BCFGL = AEFGL = BCDHL = ADEHL = FHL = ABCEFHL$
$= ABGHL = CEGHL = ACDFGHL = BDEFGHL = BDJL = ACDEJL$
$= CFJL = ABEFJL = ABCGJL = EGJL = ADFGJL = BCDEFGJL$
$= AHJL = BCEHJL = ABCDFHJL = DEFHJL = CDGHJL$
$= ABDEGHJL = BFGHJL = ACEFGHJL = BCKL = AEKL = DFKL$
$= ABCDEFKL = ABDGKL = CDEGKL = ACFGKL = BEFGKL$
$= ACDHKL = BDEHKL = ABFHKL = CEFHKL = GHKL$
$= ABCEGHKL = BCDFGHKL = ADEFGHKL = ADJKL = BCDEJKL$
$= ABCFJKL = EFJKL = CGJKL = ABEGJKL = BDFGJKL$
$= ACDEFGJKL = BHJKL = ACEHJKL = CDFHJKL = ABDEFHJKL$
$= ABCDGHJKL = DEGHJKL = AFGHJKL = BCEFGHJKL$

别名

$A = FJ = BK = CL$	$J = DE = AF = BG = CH$
$B = GJ = AK = EL$	$K = AB = CE = FG = DH$
$C = HJ = EK = AL$	$L = AC = BE = DG = FH$
$D = EJ = HK = GL$	$AD = EF = CG = BH$
$E = DJ = CK = BL$	$AE = BC = DF = GH = KL$
$F = AJ = GK = HL$	$AG = CD = BF = EH = JK$
$G = BJ = FK = DL$	$AH = BD = CF = EG = JL$
$H = CJ = DK = FL$	

8 个区组中的 2 个: $AE = BC = DF = GH = KL$

11 个因子的设计 (续)

(v) 2^{11-6} 分式析因设计　　　　分辨度 IV

设计生成元

$F = ABC \quad G = BCD \quad H = CDE \quad J = ACD \quad K = ADE \quad L = BDE$

定义关系 : $I = ABCF = BCDG = ADFG = CDEH = ABDEFH = BEGH = ACEFGH$
$= ACDJ = BDFJ = ABGJ = CFGJ = AEHJ = BCEFHJ = ABCDEGHJ$
$= DEFGHJ = ADEK = BCDEFK = ABCEGK = EFGK = ACHK$
$= BFHK = ABDGHK = CDFGHK = CEJK = ABEFJK = BDEGJK$
$= ACDEFGJK = DHJK = ABCDFHJK = BCGHJK = AFGHJK$
$= BDEL = ACDEFL = CEGL = ABEFGL = BCHL = AFHL = DGHL$
$= ABCDFGHL = ABCEJL = EFJL = ADEGJL = BCDEFGJL$
$= ABDHJL = CDFHJL = ACGHJL = BFGHJL = ABKL = CFKL$
$= ACDGKL = BDFGKL = ABCDEHKL = DEFHKL = AEGHKL$

(续)

$=BCEFGHKL=BCDJKL=ADFJKL=GJKL=ABCFGJKL$
$=BEHJKL=ACEFHJKL=CDEGHJKL=ABDEFGHJKL$

别名

$A=BCF=DFG=CDJ=BGJ=EHJ=DEK=CHK=FHL=BKL$
$B=ACF=CDG=EGH=DFJ=AGJ=FHK=DEL=CHL=AKL$
$C=ABF=BDG=DEH=ADJ=FGJ=AHK=EJK=EGL=BHL=FKL$
$D=BCG=AFG=CEH=ACJ=BFJ=AEK=HJK=BEL=GHL$
$E=CDH=BGH=AHJ=ADK=FGK=GJK=BDL=CGL=FJL$
$F=ABC=ADG=BDJ=CGJ=EGK=BHK=AHL=EJL=CKL$
$G=BCD=ADF=BEH=ABJ=CFJ=EFK=CEL=DHL=JKL$
$H=CDE=BEG=AEJ=ACK=BFK=DJK=BCL=AFL=DGL$
$J=ACD=BDF=ABG=CFG=AEH=CEK=DHK=EFL=GKL$
$K=ADE=EFG=ACH=BFH=CEJ=DHJ=ABL=CFL=GJL$
$L=BDE=CEG=BCH=AFH=DGH=EFJ=ABK=CFK=GJK$

$AB=CF=GJ=KL$	$AE=HJ=DK$	$AH=EJ=CK=FL$	$AL=FH=BK$	$BH=EG=CL=FK$
$AC=BF=DJ=HK$	$AF=BC=DG=HL$	$AJ=CD=BG=EH$	$BD=CG=FJ=EL$	$CE=DH=JK=GL$
$AD=FG=CJ=EK$	$AG=DF=BJ$	$AK=DE=CH=BL$	$BE=GH=DL$	$EF=GK=JL$

$ABD=CDF=ACG=BFG=EFH=BCJ=AFJ=DGJ=BEK=GHK=AEL=HJL=DKL$
$ABE=CEF=DFH=AGH=EGJ=BHJ=BDK=CGK=FJK=ADL=FGL=CJL=EKL$
$ABH=DEF=AEG=CFH=BEJ=GHJ=BCK=AFK=DGK=ACL=BFL=DJL=HKL$
$ACE=BEF=ADH=FGH=DEJ=CHJ=CDK=BGK=EHK=AJK=DFL=AGL=BJL$
$AEF=BCE=DEG=BDH=CGH=FHJ=DFK=AGK=BJK=CDL=BGL=EHL=AJL$

16 个区组中的 2 个: $AB=CF=GJ=KL$　8 个区组中的 4 个: $AB=CF=GJ=KL$
$AD=FG=CJ=EK$
$BD=CG=FJ=EL$

12 个因子的设计

(w) 2^{12-8} 分式析因设计　　分辨度 III

设计生成元

$E=ABC$　$F=ABD$　$G=ACD$　$H=BCD$
$J=ABCD$　$K=AB$　$L=AC$　$M=AD$

别名

$A=HJ=BK=CL=DM$
$B=GJ=AK=EL=FM$
$C=FJ=EK=AL=GM$
$D=EJ=FK=GL=AM$
$E=DJ=CK=BL=HM$
$F=CJ=DK=HL=BM$
$G=BJ=HK=DL=CM$
$H=AJ=GK=FL=EM$
$J=DE=CF=BG=AH$
$K=AB=CE=DF=GH$
$L=AC=BE=DG=FH$
$M=AD=BF=CG=EH$
$AE=BC=FG=DH=KL=JM$
$AF=BD=EG=CH=JL=KM$
$AG=EF=CD=BH=JK=LM$

8 个区组中的 2 个: $AE=BC=FG=DH=KL=JM$

(续)

13 个因子的设计

(x) 2^{13-9} 分式析因设计　　分辨度 III

设计生成元

$E=ABC \quad F=ABD \quad G=ACD \quad H=BCD$

$J=ABCD \quad K=AB \quad L=AC \quad M=AD \quad N=BC$

别名

$A=HJ=BK=CL=DM=EN$

$B=GJ=AK=EL=FM=CN$

$C=FJ=EK=AL=GM=BN$

$D=EJ=FK=GL=AM=HN$

$E=DJ=CK=BL=HM=AN$

$F=CJ=DK=HL=BM=GN$

$G=BJ=HK=DL=CM=FN$

$H=AJ=GK=FL=EM=DN$

$J=DE=CF=BG=AH=MN$

$K=AB=CE=DF=GH=LN$

$L=AC=BE=BG=FH=KN$

$M=AD=BF=CG=EH=JN$

$N=BC=AE=FG=DH=KL=JM$

$AF=BD=EG=CH=JL=KM$

$AG=EF=CD=BH=JK=LM$

8 个区组中的 2 个: $AF=BD=EG=CH=JL=KM$

14 个因子的设计

(y) 2^{14-10} 分式析因设计　　分辨度 III

设计生成元

$E=ABC \quad F=ABD \quad G=ACD \quad H=BCD \quad J=ABCD$

$K=AB \quad L=AC \quad M=AD \quad N=BC \quad O=BD$

别名

$A=HJ=BK=CL=DM=EN=FO$

$B=GJ=AK=EL=FM=CN=DO$

$C=FJ=EK=AL=GM=BN=HO$

$D=EJ=FK=GL=AM=HN=BO$

$E=DJ=CK=BL=HM=AN=GO$

$F=CJ=DK=HL=BM=GN=AO$

$G=BJ=HK=DL=CM=FN=EO$

$H=AJ=GK=FL=EM=DN=CO$

$J=DE=CF=BG=AH=MN=LO$

$K=AB=CE=DF=GH=LN=MO$

$L=AC=BE=DG=FH=KN=JO$

$M=AD=BF=CG=EH=JN=KO$

$N=BC=AE=FG=DH=KL=JM$

$O=BD=AF=EG=CH=JL=KM$

$AG=EF=CD=BH=JK=LM=NO$

8 个区组中的 2 个: $AG=EF=CD=BH=JK=LM=NO$

(续)

15 个因子的设计
(z) 2^{15-11} 分式析因设计　　　分辨度 III
设计生成元
$E = ABC \quad F = ABD \quad G = ACD \quad H = BCD \quad J = ABCD$
$K = AB \quad L = AC \quad M = AD \quad N = BC \quad O = BD \quad P = CD$
别名
$A = HJ = BK = CL = DM = EN = FO = GP$
$B = GJ = AK = EL = FM = CN = DO = HP$
$C = FJ = EK = AL = GM = BN = HO = DP$
$D = EJ = FK = GL = AM = HN = BO = CP$
$E = DJ = CK = BL = HM = AN = GO = FP$
$F = CJ = DK = HL = BM = GN = AO = EP$
$G = BJ = HK = DL = CM = FN = EO = AP$
$H = AJ = GK = FL = EM = DN = CO = BP$
$J = DE = CF = BG = AH = MN = LO = KP$
$K = AB = CE = DF = GH = LN = MO = JP$
$L = AC = BE = DG = FH = KN = JO = MP$
$M = AD = BF = CG = EH = JN = KO = LP$
$N = BC = AE = FG = DH = KL = JM = OP$
$O = BD = AF = EG = CH = JL = KM = NP$
$P = CD = EF = AG = BH = JK = LM = NO$

索　　引

A

鞍点　354, 355, 356

B

半折叠　278
饱和分式析因设计　264
保持常量因子　11
备选分式　241, 242, 249
备择假设　27, 28, 30
比例数据　292, 480, 490
边界点设计　383
变换　24, 31, 48
标准化对照　73
标准拉丁方　114
标准正态分布　23, 36, 203
别名　239, 241, 242
泊松分布　64
不可控变量　399
不完全区组设计　98, 120, 125
部分别名　271
部分混杂　236

C

残差　14, 61, 62
残差平方和　151, 323, 343
残差平方和的纯误差分量　187
测量系统能力研究　423
超饱和设计　239, 283, 284
尺度化残差　340
尺度化预测方差　372, 378
抽检特性 (OC) 曲线　32
抽样分布　17, 23, 24
处理　1, 17, 32
处理均值的对照　85
处理效应　51, 87, 139
重复测量设计　508
纯二次项弯曲的检验　209

D

单边备择假设　28, 37, 42
单纯形格点设计　381, 383, 384
单纯形设计　367, 382, 396
单因子 ANOVA 检验统计量　56
单因子完全随机化实验　51
单因子折叠设计　267
等半径设计　371
等高线图　7, 160, 196
第 I 类错误　27, 74, 75
第 II 类错误　27, 32, 284
点图　17, 18, 50
调优　347
调整的 R^2 统计量　184
定量设计因子　68, 158, 212, 296
定性设计因子　68, 212, 296
独立随机变量　107
对数联系　483, 484, 487,
对照　32, 71, 72
对照平方和　73
多重比较　61, 71, 144
多重响应　11, 362, 404

E

二阶模型　209, 210, 354
二阶设计　367, 371
二项分布　64

F

方差　9, 15, 20
方差分量　52, 64, 107
方差分量估计　415, 417, 441
方差分析　12, 48, 50
方差散布图　372
方差算子　20, 404, 405
方差稳定化变换　64, 67, 85
非标准模型　377
非几何设计　269
非线性规划　364
非正规分式设计　281
非中心 F 分布　82

分层原理 158
分散效应 48, 85, 97
分式析因设计 6, 7, 239
分式析因族 241, 251
附加平方和法 335

G

伽玛分布 483, 484, 487
概率分布 17, 19, 20
杠杆点 342
搁置数据 491
工程设计 6
构造分式析因设计
估计量 22, 28, 59
估计缺失观测值 491
广义交互作用 233, 234, 251
广义线性模型 480, 482, 483
规范设计因子 135, 175, 350
过程稳健性研究 12, 398, 411

H

盒图 18, 19, 50
恒等联系 483
互补分式设计 241
回归变量 319, 320, 332
回归的假设检验 319
回归的帽子矩阵 332
回归和方差分析 87
回归模型 17, 48, 136
回归系数 175, 176, 214
回归显著性 48, 88, 111
混合模型 415, 423, 424
混合模型的无约束形式 427
混合模型的约束形式 462
混合水平析因设计 296
混料实验 8, 380, 480
混区设计 225, 226, 234
混杂设计 257, 258, 377

J

机理模型 1
基本设计 242, 243, 245
假设检验 14, 17, 26
检验假定 76
检验统计量 27, 28, 29
简化模型 89, 90, 111
交叉设计 117
交叉因子 450, 459, 463
交错嵌套设计 450
交互作用 3, 4, 5
交换算法 378
校正平方和 53, 117, 140
近似 F 检验 415
近似多项式 70
经验模型 2, 7, 212
均方 224, 232, 247
均值模型 51, 139, 404

K

卡方分布 24, 402, 438
可估函数 134
可加模型 107
可控变量 349, 399, 400
可重复性与可再现性 421, 444, 445
控制处理 79

L

拉丁方设计 98, 112, 115
离散概率分布 20
立方体形设计 369
连续型概率分布 19
联合置信区间 59, 60, 85
联系函数 480, 484, 487
两样本 t 检验 26, 28, 33
裂裂区设计 450, 471, 472
裂区设计 472, 473, 478
岭系统 347, 362, 363
零假设 336, 416, 495
逻辑斯谛回归 480, 484, 485

M

满意度函数 365
幂变换族 481
面中心立方 (FCC) 369
模型拟合 69, 167, 342
模型适合性检验 58, 98, 146

N

内表设计 401
拟合不足检验 319
拟合响应曲面 347, 349, 366
拟合响应曲线 134
拟合优度检验 485

P

配对比较实验　38, 40
偏 F 检验　336
平衡不完全区组设计　98, 120, 132

Q

期望均方　48, 151, 153
期望值　64, 127, 421
嵌套设计　415, 429, 450
嵌套析因设计　459
球面 CCD　369
区组化　260, 269, 373
全模型　189, 190, 335
全区　369, 464, 466
全区误差　464, 465, 469
确认实验　250, 251, 338

S

散点图　50, 68,69
筛选实验　7, 171, 259
设计的 G 效率　378
设计分辨度　239, 251, 541
设计生成元　469, 541, 534
实验策略　1, 2, 3
实验单元　11, 12, 41
实验的最佳猜测法　3
实验方法误差率　59
双边备择假设　27
随机抽样　21
随机化　26, 32, 51
随机化检验　32, 58, 141
随机化区组设计　40, 105, 111
随机化完全区组设计 (RCBD)　98
随机效应模型　52, 415, 417

T

泰勒级数　404
讨厌因子　10, 11, 12
特征向量　361
特征值　356
统计假设　27
统计检验的显著性水平　17, 29
统计量　21, 23, 28
统计模型　27, 52, 112
推断分布　42

W

外表设计　401
完全混区设计　236
完全随机化　129, 162, 216
完全随机化实验　26, 92, 98
完全折叠　267, 276, 281
伪分量　384
位置效应　85, 86, 401
稳定岭系统　362
稳健参数设计　15, 398
稳健产品设计　138
稳健过程　1, 15, 398
无计划实验　319
无偏估计量　22, 151, 323
无重复析因设计　192, 198, 201
误差传播　405
误差传递法　404

X

析因设计中的不平衡数据　480
析因实验或设计　4, 239, 296
线性统计模型　51, 162, 471
响应变量　480, 483, 503
响应曲面　7, 8, 136
响应曲面的稳定点　354, 360
响应曲面法　347
响应曲面设计　297, 348, 368
响应曲面图　136, 177, 378
像差　257
小复合设计　371, 411, 414
效应的半正态概率图　276, 277, 402
效应模型　415, 416, 428
效应稀疏原理　192
协变量　12, 493, 495
协方差　495, 496, 498
协方差分析　498, 499, 503
协方差矩阵　319, 323, 443
信噪比　103
序贯实验　239, 240, 275
旋转设计　368, 372, 375
旋转性　347, 368, 369
学生化残差　339, 340, 341
学生化极差统计量　76, 77, 536

Y

样本标准差　21, 40, 41
样本方差　42, 43, 53

样本均值 9, 21, 22
一般回归显著性检验 48, 87, 89
一次一因子实验 4, 137, 138
一阶模型 209, 210, 342
一阶设计 367, 373, 395
因变量 319
因果图 12
因子筛选 11, 12, 171
隐藏重复 195
影响诊断 319
预测变量 320
预测方差准则 378
预测区间 251
约束最大似然估计法 443
约束最优化 364

Z

噪声变量 399, 400, 401
正交对照 48, 73, 74
正交拉丁方 98, 118, 118
正交拉丁方设计 98, 117, 118
正交区组化 373
正交设计 160, 175, 284
正态分布 332, 337, 402
正态概率图 17, 30, 31
正则形式 347, 356, 361
直积表设计 399, 400, 401
指数分布 484
秩变换 48
置信系数 35
置信区间 14, 17, 34
中心点 171, 209, 211
中心复合设计 7, 210, 347
中心极限定理 23, 36, 37
轴试验 210, 367, 383
主分式设计 244, 250, 266
主区组 228, 229, 234
子区 (或裂区) 464
子区误差 464, 465, 472
自变量 319, 320, 347
自由度 22, 24, 25
组合表设计 398, 401, 414
最大似然估计法 441, 484
最速上升法 177, 347, 349
最速上升路径 349, 350, 352
最小低阶混杂设计 257
最小二乘法 69, 109, 323
最小二乘正规方程组 97, 125, 321
最小方差估计量 22
最小分辨度为 IV 的设计 282
最小显著性差异 77
最优设计 272, 273, 377

其 他

2^3 析因设计的回归分析 326
2^k 设计的标准顺序 175
2^k 设计的回归系数的标准误 189
2^k 设计的效应的标准误 189
2^k 设计投影 195
ANOVA 随机因子 51, 415
ANOVA 中的方差相等性检验 64
A 最优设计 378
Bartlett 检验 65
Box-Behnken 设计 369, 370, 394
Cochran 定理 55
Dunnett 检验 79, 92, 530
D 最优设计 272, 377, 378
Fisher LSD 方法 105, 128, 129
F 分布 25, 43, 72
G 最优准则 378
Hadamard 矩阵设计 283
Hosmer-Lemeshow 拟合优度检验 485
Kruskal-Wallis 检验法 48, 90, 97
Levene 检验 65, 66, 96
Logit 联系 484
Ockham 剃刀原则 244
Pearson 拟合优度检验 485
Plackett-Burman 设计 269, 270, 271
PRESS 统计量 184, 224, 341
PRESS 残差 340
Tukey 检验 76, 77, 94
t 分布 71, 86, 166
V 最优设计 378
Yates 顺序 175